The Capacity Region of a Class of 3-Receiver Broadcast Channels with Degraded Message Sets

Chandra Nair, *Member, IEEE,* and Abbas El Gamal, *Fellow, IEEE*

Abstract

Körner and Marton established the capacity region for the 2-receiver broadcast channel with degraded message sets. Recent results and conjectures suggest that a straightforward extension of the Körner-Marton region to more than 2 receivers is optimal. This paper shows that this is not the case. We establish the capacity region for a class of 3-receiver broadcast channels with 2 degraded message sets and show that it can be strictly larger than the straightforward extension of the Körner-Marton region. The key new idea is indirect decoding, whereby a receiver who cannot directly decode a cloud center, finds it indirectly by decoding satellite codewords. This idea is then used to establish new inner and outer bounds on the capacity region of the general 3-receiver broadcast channel with 2 and 3 degraded message sets. We show that these bounds are tight for some nontrivial cases. The results suggest that the capacity of the 3-receiver broadcast channel with degraded message sets is as at least as hard to find as the capacity of the general 2-receiver broadcast channel with common and private message.

I. Introduction

A broadcast channel with degraded message sets represents a scenario where a sender wishes to communicate a common message to *all* receivers, a first private message to a first subset of the receivers, a second private message to a second subset of the first subset and so on. Such scenario can arise, for example, in video or music broadcasting over a wireless network to nested subsets of receivers at varying levels of quality. The common message represents the lowest quality version to be sent to all receivers, the first private message represents the additional information needed for the first subset of receivers to decode the second lowest quality version, and so on. What is the set of simultaneously achievable rates for communicating such degraded message sets over the network?

This question was first studied by Körner and Marton in 1977 [1]. They considered a general 2-receiver discrete-memoryless broadcast channel with sender X and receivers Y_1 and Y_2. A common message $M_0 \in [1, 2^{nR_0}]$ is to be sent to both receivers and a private message $M_1 \in [1, 2^{nR_1}]$ is to be sent only to receiver Y_1. They showed that the capacity region is given by the set of all rate pairs (R_0, R_1) such that [1]

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_1), I(U;Y_2)\}, \\ R_1 &\leq I(X;Y_1|U), \end{aligned} \tag{1}$$

for some $p(u,x)$. These rates are achieved using superposition coding [2]. The common message is represented by the auxiliary random variable U and the private message is superimposed to generate X. The main contribution of [1] is proving a strong converse using the technique of images-of-a-set [3].

Extending the Körner-Marton result to more than 2 receivers has remained open even for the simple case of 3 receivers Y_1, Y_2, Y_3 with 2 degraded message sets, where a common message M_0 is to be sent to all receivers and a private message M_1 is to be sent only to receiver Y_1. The straightforward extension of the Körner-Marton region to this case yields the achievable rate region consisting of the set of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_1), I(U;Y_2), I(U;Y_3)\}, \\ R_1 &\leq I(X;Y_1|U), \end{aligned} \tag{2}$$

Chandra Nair was partly supported by the Direct Grant for research at the Chinese University of Hong Kong

[1]The Körner-Marton characterization does not include the second term inside the min in the first inequality, $I(U;Y_1)$. Instead it includes the bound $R_1 + R_2 \leq I(X;Y_1)$. It can be easily shown that the two characterizations are equivalent.

for some $p(u, x)$. Is this region optimal?

In [4], it was shown that the above region (and its natural extension to $k > 3$ receivers) is optimal for a class of product discrete-memoryless and Gaussian broadcast channels, where each of the receivers who decode only the common message is a degraded version of the unique receiver that also decodes the private message. In [5], it was shown that a straightforward extension of Körner-Marton region is optimal for the class of linear deterministic broadcast channels, where the operations are performed in a finite field. In addition to establishing the degraded message set capacity for this class the authors gave an explicit characterization of the optimal auxiliary random variables. In a recent paper Borade et al. [6] introduced *multilevel* broadcast channels, which combine aspects of degraded broadcast channels and broadcast channels with degraded message sets. They established an achievable rate region as well as a "mirror-image" outer bound for these channels. Their achievable rate region is again a straightforward extension of the Körner-Marton region to k-receiver multilevel broadcast channels. In particular, Conjecture 5 of [6] states that the capacity region for the 3-receiver multilevel broadcast channels depicted in Figure 1 is the set of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_2), I(U;Y_3)\}, \\ R_1 &\leq I(X;Y_1|U), \end{aligned} \tag{3}$$

for some $p(u, x)$. Note that this region, henceforth referred to as *the BZT region*, is the same as (2) because in the multilevel broadcast channel Y_3 is a degraded version of Y_1 and therefore $I(U;Y_3) \leq I(U;Y_1)$.

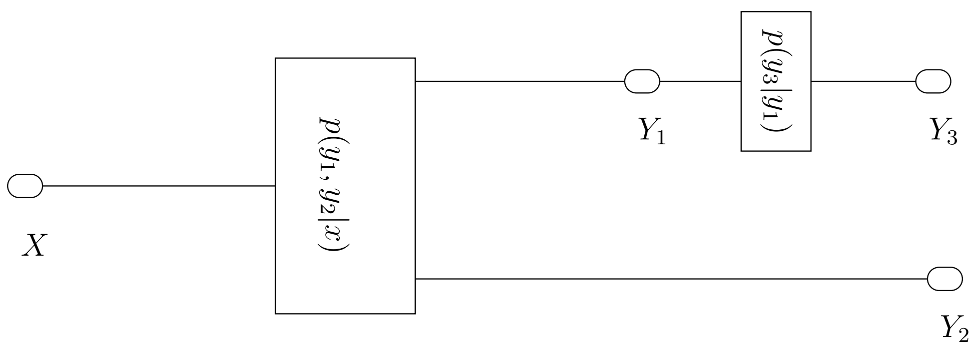

Fig. 1. Multilevel 3-receiver broadcast channels. Message M_0 is to be sent to all receivers and message M_1 is to be sent only to Y_1.

In this paper we show that the straightforward extension of the Körner-Marton region to more than 2 receivers is not in general optimal. We establish the capacity region of the multilevel broadcast channels depicted in Figure 1 as the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U_2;Y_2), I(U_1;Y_3)\}, \\ R_0 + R_1 &\leq \min\{I(U_1;Y_3) + I(X;Y_1|U_1), I(U_2;Y_2) + I(X;Y_1|U_2)\}, \end{aligned}$$

for some $p(u_1)p(u_2|u_1)p(x|u_2)$, and show that it can be strictly larger than the BZT region. In our coding scheme, the common message M_0 is represented by U_1 (the cloud centers), part of M_1 is superimposed on U_1 to obtain U_2 (satellite codewords), and the rest of M_1 is superimposed on U_2 to yield X. Receivers Y_1 and Y_3 find M_0 by decoding U_1, whereas receiver Y_2 who may not be able to directly decode U_1, finds M_0 *indirectly* by decoding a list of satellite codewords. After decoding U_1, receiver Y_1 finds M_1 by proceeding to decode U_2 then X.

The rest of the paper is organized as follows. In Section II, we provide needed definitions. In Section III, we establish the capacity region for the multilevel broadcast channel in Figure 1 (Theorem 1). In Section IV, we show through an example that the capacity region for the multilevel broadcast channel can be strictly larger than the BZT region. In Section V, we extend the results on the multilevel broadcast channel to establish inner and outer bounds on the capacity region of the general 3-receiver broadcast

channel with 2 degraded message sets (Propositions 5 and 6). We show that these bounds are tight when Y_1 is *less noisy* than Y_3 (Proposition 7), which is a more relaxed condition than the degradedness condition of the multilevel broadcast channel model. We then extend the inner bound to 3 degraded message sets (Theorem 2). Although Proposition 5 is a special case of Theorem 2, it is presented earlier for clarity of exposition. Finally, we show that the inner bound of Theorem 2 when specialized to the case of 2 degraded message sets, where M_0 is to be sent to all receivers and M_1 is to be sent to Y_1 and Y_2, reduces to the straightforward extension of the Körner-Marton region (Corollary 1). We show that this inner bound is tight for deterministic broadcast channels (Proposition 8) and when Y_1 is less noisy than Y_3 and Y_2 is less noisy than Y_3 (Proposition 9). Finally, we outline a general approach to obtaining inner bounds on capacity for general k-receiver broadcast channel scenarios that uses the new idea of indirect decoding.

II. Definitions

Consider a discrete-memoryless 3-receiver broadcast channel consisting of an input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}_1$, $\mathcal{Y}_2$ and $\mathcal{Y}_3$, and a probability transition function $p(y_1, y_2, y_3|x)$.

A $(2^{nR_0}, 2^{nR_1}, n)$ *2-degraded message set* code for a 3-receiver broadcast channel consists of (i) a pair of messages (M_0, M_1) uniformly distributed over $[1, 2^{nR_0}] \times [1, 2^{nR_1}]$, (ii) an encoder that assigns a codeword $x^n(m_0, m_1)$, for each message pair $(m_0, m_1) \in [1, 2^{nR_0}] \times [1, 2^{nR_1}]$, and (iii) three decoders, one that maps each received y_1^n sequence into an estimate $(\hat{m}_{01}, \hat{m}_1) \in [1, 2^{nR_0}] \times [1, 2^{nR_1}]$, a second that maps each received y_2^n sequence into an estimate $\hat{m}_{02} \in [1, 2^{nR_0}]$, and a third that maps each received y_3^n sequence into an estimate $\hat{m}_{03} \in [1, 2^{nR_0}]$.

The probability of error is defined as

$$P_e^{(n)} = \mathrm{P}\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_{0k} \neq M_0 \text{ for } k = 1, 2, \text{ or } 3\}.$$

A rate tuple (R_0, R_1) is said to be achievable if there exists a sequence of $(2^{nR_0}, 2^{nR_1}, n)$ 2-degraded message set codes with $P_e^{(n)} \to 0$. The capacity region of the broadcast channel is the closure of the set of achievable rates.

A 3-receiver *multilevel* broadcast channel [6] is a 3-receiver broadcast channel with 2 degraded message sets where $p(y_1, y_2, y_3|x) = p(y_1, y_2|x)p(y_3|y_1)$ for every $(x, y_1, y_2, y_3) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ (see Figure 1).

In addition to considering the multilevel 3-receiver broadcast channel and the general 3-receiver broadcast channel with 2 degraded message sets, we shall also consider the following two scenarios:

1) 3-receiver broadcast channel with 3 message sets, where M_0 is to be sent to all receivers, M_1 is to be sent to Y_1 and Y_2, and M_2 is to be sent only to Y_1.
2) 3-receiver broadcast channel with 2 degraded message sets, where M_0 is to be sent all receivers and M_1 is to be sent to both Y_1 and Y_2.

Definitions of codes, achievability and capacity regions for these cases are straightforward extensions of the above definitions. Clearly, the 2 degraded message set scenarios are special cases of the 3 degraded message set. Nevertheless, we shall begin with the special class of multilevel broadcast channel because we are able to establish its capacity region.

III. Capacity of 3-Receiver Multilevel Broadcast Channel

The key result of this paper is given in the following theorem.

Theorem 1: The capacity region of the 3-receiver multilevel broadcast channel in Figure 1 is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_0 &\leq \min\{I(U_1; Y_3), I(U_2; Y_2)\}, \\ R_0 + R_1 &\leq \min\{I(U_1; Y_3) + I(X; Y_1|U_1), I(U_2; Y_2) + I(X; Y_1|U_2), \}. \end{aligned} \tag{4}$$

for some $p(u_1)p(u_2|u_1)p(x|u_2)$, where the cardinalities of the auxiliary random variables satisfy $\|\mathcal{U}_1\| \leq \|\mathcal{X}\| + 4$ and $\|U_2\| \leq \|\mathcal{X}\|^2 + 5\|\mathcal{X}\| + 4$.

Remarks:

1) It is easy to show by setting $U_1 = U_2 = U$ in the above theorem that the BZT region (3) is contained in the capacity region (4). We show in the next section that the capacity region (4) can be strictly larger the BZT region.
2) It is straightforward to show that the above region is convex and therefore there is no need to use a time-sharing auxiliary random variable.

The proof of Theorem 1 is given in the following subsections.

A. Converse

We show that the region in Theorem 1 forms an outer bound to the capacity region. The proof is quite similar to previous weak converse and outer bound proofs for 2-receiver broadcast channels (e.g., see [7], [8], [9]). Suppose we are given a sequence of codes for the multilevel broadcast channel with $P_e^{(n)} \to 0$. For each code, we form the empirical distribution for M_0, M_1, X^n.

Since $X \to Y_1 \to Y_3$ forms a *physically degraded* broadcast channel, it follows that the code rate pair (R_0, R_1) must satisfy the inequalities

$$\begin{aligned} R_0 &\leq I(U_1; Y_3), \\ R_1 &\leq I(X; Y_1 | U_1), \end{aligned}$$

for some $p(u_1, x)$ [7], where U_1, X, Y_1, Y_3 are defined as follows. Let $U_{1i} = (M_0, Y_1^{i-1}),\ i = 1, \ldots, n$, and let Q be a time-sharing random variable uniformly distributed over the set $\{1, 2, ..., n\}$ and independent of X^n, Y_1^n, Y_2^n, Y_3^n. We then set $U_1 = (Q, U_{1Q})$ and $X = X_Q$, $Y_1 = Y_{1Q}$, and $Y_3 = Y_{3Q}$. Thus, we have shown that

$$\begin{aligned} R_0 &\leq I(U_1; Y_3), \\ R_0 + R_1 &\leq I(U_1; Y_3) + I(X; Y_1 | U_1). \end{aligned}$$

Next, since the decoding requirements of the broadcast channel $X \to (Y_1, Y_2)$ makes it a broadcast channel with degraded message sets, the code rate pair must satisfy the inequalities

$$\begin{aligned} R_0 &\leq \min\{I(U_2; Y_2), I(U_2, Y_1)\}, \\ R_0 + R_1 &\leq I(U_2; Y_2) + I(X; Y_1 | U_2), \end{aligned}$$

for some $p(u_2, x)$ [8], where U_2 is identified as follows. Let $U_{2i} = (M_0, Y_1^{i-1}, Y_2{}_{i+1}^{n}),\ i = 1, \ldots, n$, then we set $U_2 = (Q, U_{2Q})$.

Combining the above two outer bounds, we see that $U_1 \to U_2 \to X$ forms a Markov chain. Observe that this Markov nature of the auxiliary random variables along with the degraded nature of $X \to Y_1 \to Y_3$ implies that $I(U_2; Y_1) \geq I(U_2; Y_3) \geq I(U_1; Y_3)$.

Thus we have shown that the code rate pair (R_0, R_1) must satisfy the inequalities

$$\begin{aligned} R_0 &\leq \min\{I(U_1; Y_3), I(U_2; Y_2)\}, \\ R_0 + R_1 &\leq \min\{I(U_1; Y_3) + I(X; Y_1 | U_1), I(U_2; Y_2) + I(X; Y_1 | U_2)\}, \end{aligned}$$

for some $p(u_1)p(u_2|u_1)p(x|u_2)$. This establishes the converse to Theorem 1.

B. Achievability

The interesting part of the proof of Theorem 1 is achievability. Specifically, step 3 of the decoding procedure for Case 2 below describes a key contribution of this paper. We show how Y_2 can find M_0 without directly decoding U_1^n or uniquely decoding U_2^n.

To show achievability of any rate pair (R_0, R_1) in region (4), because of its convexity, it suffices to show the achievability of any rate pair (R_0, R_1) such that

$$R_0 = \min\{I(U_1;Y_3), I(U_2;Y_2)\} - \delta$$
$$R_0 + R_1 = \min\{I(U_1;Y_3) + I(X;Y_1|U_1), I(U_2,Y_2) + I(X;Y_1|U_2)\} - 3\delta,$$

for some $U_1 \to U_2 \to X$ and any $\delta > 0$.

Rewriting the second inequality we obtain

$$R_0 + R_1 = I(U_1;Y_3) + \min\{I(U_2;Y_1|U_1), I(U_2;Y_3) - I(U_1;Y_3)\} + I(X;Y_1|U_2) - 3\delta.$$

Now consider the following two cases:

Case 1: $I(U_1;Y_3) > I(U_2;Y_2)$: The rates reduce to

$$R_0 = I(U_2;Y_2) - \delta$$
$$R_1 = I(X;Y_1|U_2) - 2\delta.$$

This pair can be achieved via a simple superposition coding scheme [2].

Case 2: $I(U_1;Y_3) \le I(U_2;Y_2)$: The rates reduce to

$$R_0 = I(U_1;Y_3) - \delta$$
$$R_1 = I(X;Y_1|U_2) + \min\{I(U_2;Y_1|U_1), I(U_2,Y_2) - I(U_1;Y_3)\} - 2\delta.$$

Let $S_1 = \min\{I(U_2;Y_1|U_1), I(U_2,Y_2) - I(U_1;Y_3)\} - \delta$ and $S_2 = I(X;Y_1|U_2) - \delta$, then $R_1 = S_1 + S_2$.

Code Generation:

Fix $p(u_1)p(u_2|u_1)p(x|u_2)$ that satisfies the condition of Case 2. Generate $2^{nR_0} = 2^{n(I(U_1;Y_3)-\delta)}$ sequences $U_1^n(1), \ldots, U_1^n(2^{nR_0})$ distributed uniformly at random over the set of ϵ-typical† U_1^n sequences, where $\delta \to 0$ as $\epsilon \to 0$. For each $U_1^n(m_0)$, generate $2^{nS_1} = 2^{n(\min\{I(U_2;Y_1|U_1), I(U_2,Y_2)-I(U_1;Y_3)\}-\delta)}$ sequences $U_2^n(m_0,1)$, $U_2^n(m_0,2)$, ..., $U_2^n(m_0,2^{nS_1})$ distributed uniformly at random over the set of conditionally ϵ-typical U_2^n sequences. For each $U_2^n(m_0,s_1)$ generate $2^{nS_2} = 2^{n(I(X;Y_1|U_2)-\delta)}$ sequences $X^n(m_0,s_1,1)$, $X^n(m_0,s_1,2)$, ..., $X^n(m_0,s_1,2^{nS_2})$ distributed uniformly at random over the set of conditionally ϵ-typical X^n sequences.

Encoding:

To send the message pair $(m_0,m_1) \in [1,2^{nR_0}] \times [1,2^{nR_1}]$, the sender expresses m_1 by the pair $(s_1,s_2) \in [1,2^{nS_1}] \times [1,2^{nS_2}]$ and sends $X^n(m_0,s_1,s_2)$.

Decoding and Analysis of Error Probability:

1) Receiver Y_3 declares that m_0 is sent if it is the unique message such that $U_1^n(m_0)$ and Y_3^n are jointly ϵ-typical. It is easy to see that this can be achieved with arbitrarily small probability of error because $R_0 = I(U_1;Y_3) - \delta$.
2) Receiver Y_1 first declares that m_0 is sent if it is the unique message such that $U_1^n(m_0)$ and Y_1^n are jointly ϵ-typical. This decoding step succeeds with arbitrarily high probability because $R_0 = I(U_1;Y_3) - \delta \le I(U_1;Y_1) - \delta$. It then declares that s_1 is sent if it is the unique index such that $U_2^n(m_0,s_1)$ and Y_1^n are jointly ϵ-typical. This decoding step succeeds with arbitrarily high probability because $S_1 \le I(U_2;Y_1|U_1) - \delta$. Finally it declares that s_2 is sent if it is the unique index such that $X^n(m_0,s_1,s_2)$ and Y_1^n are jointly ϵ-typical. This decoding step succeeds with high probability because $S_2 = I(X;Y_1|U_2) - \delta$.
3) Receiver Y_2 finds M_0 as follows. It declares that $m_0 \in [1,2^{nR_0}]$ is sent if it is the unique index such that $U_2^n(m_0,s_1)$ and Y_2^n are jointly ϵ-typical for some $s_1 \in [1,2^{nS_1}]$. Suppose $(1,1) \in [1,2^{nR_0}] \times$

†We assume strong typicality throughout this paper [10].

$[1, 2^{nS_1}]$ is the message pair sent, then the probability of error averaged over the choice of codebooks can be upper bounded as follows

$$\begin{aligned}
P_e^{(n)} &\le \mathrm{P}\{(U_2^n(1,1), Y_2^n) \text{ are not jointly } \epsilon\text{-typical}\} \\
&\qquad + \mathrm{P}\{(U_2^n(m_0,s_1), Y_2^n) \text{ are jointly } \epsilon\text{-typical for some } m_0 \neq 1\} \\
&\overset{(a)}{<} \delta' + 2^{n(R_0+S_1)} \sum_{m_0 \neq 1} \sum_{s_1} \mathrm{P}\{(U_2^n(m_0,s_1), Y_2^n) \text{ jointly } \epsilon\text{-typical}\} \\
&\overset{(b)}{\le} \delta' + 2^{n(R_0+S_1)} 2^{-n(I(U_2;Y_2)-\delta)} \\
&\overset{(c)}{\le} \delta' + 2^{-n\delta},
\end{aligned}$$

where (a) follows by the union of events bound, (b) follows by the fact that for $m_0 \neq 1$, $U_2^n(m_0, s_1)$ and Y_2^n are generated completely independently and thus each probability term under the sum is upper bounded by $2^{-n(I(U_2;Y_2)-\delta)}$ [10], (c) follows because by construction $R_0+S_1 \le I(U_2;Y_2)-2\delta$, $\delta' \to 0$ as $\epsilon \to 0$. Thus with arbitrarily high probability, any jointly ϵ-typical $U_2^n(m_0, s_1)$ with the received Y_2^n sequence must be of the form $U_2^n(1, s_1)$, and receiver Y_2 can correctly decodes M_0 with arbitrarily small probability of error. Note that Y_2 may or may not be able to uniquely decode $U_2^n(1,1)$. However, it finds the correct common message with arbitrarily small probability of error even if its rate $R_0 > I(U_1;Y_2)$!

Thus all receivers can decode their intended messages with arbitrarily small probability of error and hence the rate pair $R_0 = I(U_1;Y_3) - \delta, R_1 = I(X;Y_1|U_2) + \min\{I(U_2;Y_1|U_1), I(U_2,Y_2) - I(U_1;Y_3)\} - 2\delta$ is achievable. This completes the proof of achievability of Theorem 1.

Remarks:

1) We denote the decoding technique used in step 3 as *indirect decoding*, since Y_2 decodes the cloud center U_1 indirectly by decoding satellite codewords.
2) There is no need to break up the coding scheme into two cases. The coding scheme for Case 2 suffices. This will become clear when we prove the achievable region for the general case of 3-receivers with 2 degraded message sets in Proposition 5.

C. Bounds on Cardinality

Using the strengthened Carathéodory theorem by Fenchel and Eggleston [11] it can be readily shown that for any choice of the auxiliary random variable U_1, there exists a random variable V_1 with cardinality bounded by $\|\mathcal{X}\| + 1$ such that $I(U_1;Y_3) = I(V_1;Y_3)$ and $I(X;Y_1|U_1) = I(X;Y_1|V_1)$. Similarly for any choice of U_2, one can obtain a random variable V_2 with cardinality bounded by $\|\mathcal{X}\| + 1$ such that $I(U_2;Y_2) = I(V_2;Y_2)$ and $I(X;Y_1|U_2) = I(X;Y_1|V_2)$. While this preserves the region, it is not clear that the new random variables V_1, V_2 would preserve the Markov condition $V_1 \to V_2 \to X$. To circumvent this problem we adapt arguments from [11] to establish the cardinality bounds stated in Theorem 1. For completeness, we provide an outline of the argument.

Given $U_1 \to U_2 \to X \to (Y_1, Y_2, Y_3)$, we need to show the existence of random variables V_1, V_2 such that the following conditions hold: $V_1 \to V_2 \to X$ forms a Markov chain, $I(V_1;Y_3) = I(U_1;Y_3)$, $I(V_2;Y_2) = I(U_2;Y_2)$, $I(X;Y_1|V_1) = I(X;Y_1|V_1)$, and $I(X;Y_1|V_2) = I(X;Y_1|U_2)$. Further, the cardinalities of the new random variables must satisfy $\|\mathcal{V}_1\| \le \|\mathcal{X}\| + 4$, $\|\mathcal{V}_2\| \le \|\mathcal{X}\|^2 + 5\|\mathcal{X}\| + 4$.

This argument is proved in two steps. In the first step a random variable V_1 and transition probabilities $p(u_2|v_1)$ are constructed such that the following are held constant: $p(x)$, the marginal probability of X (and hence Y_1, Y_2, Y_3), $H(Y_1|U_1)$, $H(Y_2|U_1)$, $H(Y_3|U_1)$, $H(Y_2|U_2, U_1)$, and $H(Y_1|U_2, U_1)$. Using standard arguments [12], [11], there exists a random variable V_1 and transition probabilities $p(u_2|v_1)$, with cardinality of V_1 bounded by $\|\mathcal{X}\| + 4$, such that the above equalities are achieved. In particular the marginals

of X, Y_1, Y_2, Y_3 are held constant. However the distribution of U_2 is not necessarily held constant and hence we shall denote the resulting random variable as U_2'.

We thus have random variables $V_1 \to U_2' \to X$ such that

$$\begin{aligned} I(V_1;Y_3) &= I(U_1;Y_3), \\ I(V_1;Y_2) &= I(U_1;Y_2), \\ I(X;Y_1|V_1) &= I(X;Y_1|U_1), \\ I(U_2';Y_1|V_1) &= I(U_2;Y_1|U_1). \end{aligned} \tag{5}$$

In the second step, for each $V_1 = v_1$ a new random variable $V_2(v_1)$ is found such that the following are held constant: $p(x|v_1)$, the marginal distribution of X conditioned on $V_1 = v_1$, $H(Y_1|U_2', V_1 = v_1)$, and $H(Y_2|U_2', V_1 = v_1)$. Again standard arguments imply that there exists a random variable $V_2(v_1)$ and transition probabilities $p(x|v_2(v_1))$, with cardinality of V_1 bounded by $\|\mathcal{X}\| + 1$, such that the above equalities are achieved. This in particular implies that

$$\begin{aligned} I(V_2(V_1);Y_2|V_1) &= I(U_2';Y_2|V_1) = I(U_2;Y_2|U_1), \\ I(V_2(V_1);Y_1|V_1) &= I(U_2';Y_1|V_1) = I(U_2;Y_1|U_1). \end{aligned} \tag{6}$$

Now, set $V_2 = (V_1, V_2(V_1))$ and observe the following as a consequence of Equations (5) and (6).

$$\begin{aligned} I(V_2;Y_2) &= I(V_1;Y_2) + I(V_2(V_1);Y_2|V_1) = I(U_1;Y_2) + I(U_2;Y_2|U_1) = I(U_2;Y_2), \\ I(X;Y_1|V_2) &= I(X;Y_1|V_1) - I(V_2(V_1);Y_1|V_1) = I(X;Y_1|U_1) - I(U_2;Y_1|U_1) = I(X:Y_1|U_2). \end{aligned}$$

We thus have the required random variables V_1, V_2 satisfying the cardinality bounds $\|\mathcal{X}\| + 4$ and $(\|\mathcal{X}\| + 4)(\|\mathcal{X}\| + 1)$, respectively as desired.

IV. Multilevel Product Broadcast Channel

In this section we show that the BZT region can be strictly smaller than the capacity region in Theorem 1. Consider the product of 3-receiver broadcast channels given by the Markov relationships

$$\begin{aligned} &X_1 \to Y_{21} \to Y_{11} \to Y_{31}, \\ &X_2 \to Y_{12} \to Y_{32}. \end{aligned} \tag{7}$$

In Appendix I we derive the following simplified characterizations for the capacity and the BZT regions.

Proposition 1: The BZT region for the above product channel reduces to the set of rate pairs (R_0, R_1) such that

$$R_0 \le I(V_1;Y_{31}) + I(V_2;Y_{32}), \tag{8a}$$
$$R_0 \le I(V_1;Y_{21}), \tag{8b}$$
$$R_1 \le I(X_1;Y_{11}|V_1) + I(X_2;Y_{12}|V_2), \tag{8c}$$

for some $p(v_1)p(v_2)p(x_1|v_1)p(x_2|v_2)$.

Proposition 2: The capacity region for the product channel reduces to the set of rate pairs (R_0, R_1) such that

$$R_0 \le I(V_{11};Y_{31}) + I(V_{12};Y_{32}), \tag{9a}$$
$$R_0 \le I(V_{21};Y_{21}), \tag{9b}$$
$$R_0 + R_1 \le I(V_{11};Y_{31}) + I(V_{12};Y_{32}) + I(X_1;Y_{11}|V_{11}) + I(X_2;Y_{12}|V_{12}), \tag{9c}$$
$$R_0 + R_1 \le I(V_{21};Y_{21}) + I(X_1;Y_{11}|V_{21}) + I(X_2;Y_{12}|V_{12}), \tag{9d}$$

for some $p(v_{11})p(v_{21}|v_{11})p(x_1|v_{21})p(v_{12})p(x_2|v_{12})$.

Now we compare these two regions via the following example.

Example:

Consider the multilevel product broadcast channel example in Figure 2, where: $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_{12} = \mathcal{Y}_{21} = \{0,1\}$, and $\mathcal{Y}_{11} = \mathcal{Y}_{31} = \mathcal{Y}_{32} = \{0,E,1\}$, $Y_{21} = X_1$, $Y_{12} = X_2$, the channels $Y_{21} \to Y_{11}$ and $Y_{12} \to Y_{32}$ are binary erasure channels (BEC) with erasure probability $\frac{1}{2}$, and the channel $Y_{11} \to Y_{31}$ is given by the transition probabilities: $\mathrm{P}\{Y_{31} = E | Y_{11} = E\} = 1,\ \mathrm{P}\{Y_{31} = E | Y_{11} = 0\} = \mathrm{P}\{Y_{31} = E | Y_{11} = 1\} = 2/3,\ \mathrm{P}(Y_{31} = 0 | Y_{11} = 0\} = \mathrm{P}\{Y_{31} = 1 | Y_{11} = 1\} = 1/3$. Therefore, the channel $X_1 \to Y_{31}$ is effectively a BEC with erasure probability $5/6$.

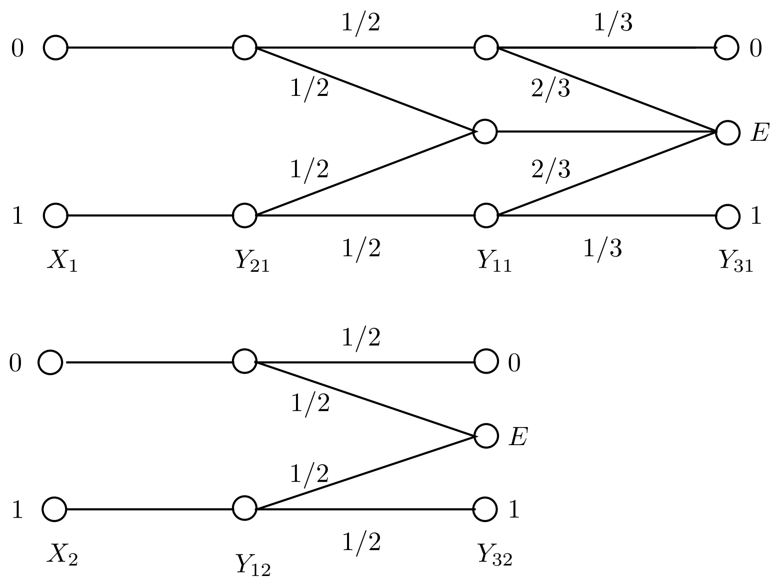

Fig. 2. Product multilevel broadcast channel example.

The BZT region can be simplified to the following.

Proposition 3: The BZT region for the above example reduces to the set of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &\le \min\left\{\frac{p}{6} + \frac{q}{2}, p\right\}, \\ R_1 &\le \frac{1-p}{2} + 1 - q. \end{aligned} \tag{10}$$

for some $0 \le p, q \le 1$.

The proof of this proposition is given in Appendix I. It is quite straightforward to see that $(R_0, R_1) = (\frac{1}{2}, \frac{5}{12})$ lies on the boundary of this region.

The capacity region can be simplified to the following

Proposition 4: The capacity region for the channel in Figure 2 reduces to set of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &\le \min\left\{\frac{r}{6} + \frac{s}{2}, t\right\}, \\ R_0 + R_1 &\le \min\left\{\frac{r}{6} + \frac{s}{2} + \frac{1-r}{2} + 1 - s, t + \frac{1-t}{2} + 1 - s\right\}, \end{aligned} \tag{11}$$

for some $0 \le r \le t \le 1, 0 \le s \le 1$.

The proof of this proposition is also given in Appendix I. Note that substituting $r = t$ yields the BZT region. By setting $r = 0, s = 1, t = 1$ it is easy to see that $(R_0, R_1) = (1/2, 1/2)$ lies on the boundary of the capacity region. On the other hand, for $R_0 = 1/2$, the maximum achievable R_1 in the BZT region is $5/12$. Thus the capacity region is strictly larger than the BZT region.

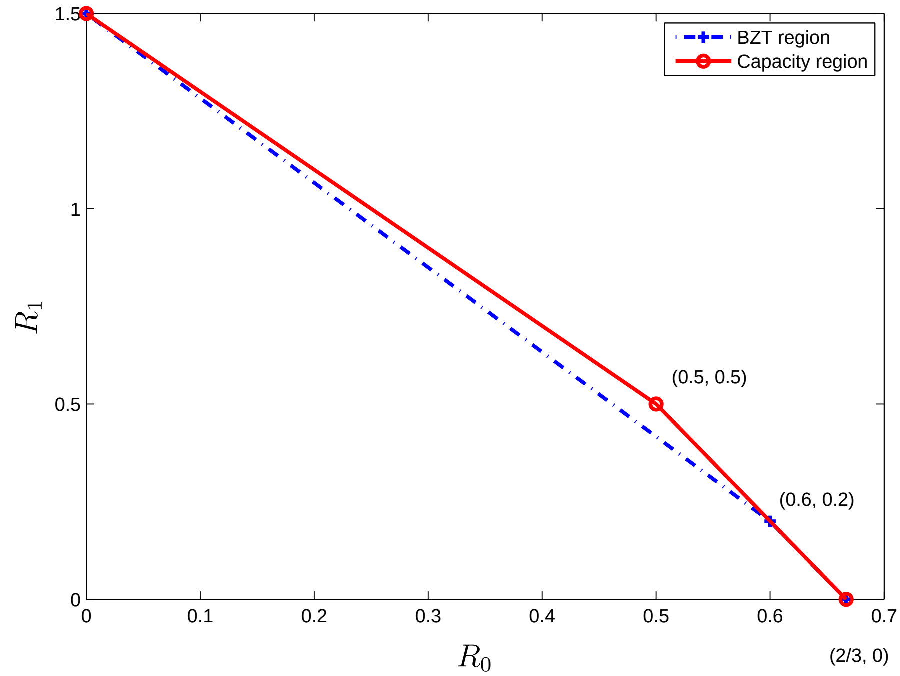

Fig. 3. The BZT and the capacity regions for the channel in Figure 2.

Figure 3 plots the BZT region and the capacity region for the example channel. Both regions are specified by two line segments. The boundary of the BZT regions consists of the line segments: $(0, 3/2)$ to $(0.6, 0.2)$ and $(0.6, 0.2)$ to $(2/3, 0)$. The capacity region on the other hand is formed by the pair of line segments: $(0, 3/2)$ to $(1/2, 1/2)$ and $(1/2, 1/2)$ to $(2/3, 0)$. Note that the boundaries of the two regions coincide on the line segment joining $(0.6, 0.2)$ to $(2/3, 0)$.

Remarks:

1) Consider a 3-receiver Gaussian product multilevel broadcast channel, where

$$Y_{21} = X_1 + Z_1,\ Y_{11} = Y_{21} + Z_2,\ Y_{31} = Y_{11} + Z_3,$$
$$Y_{22} = X_2 + Z_4,\ Y_{32} = Y_{22} + Z_5.$$

The noise power for Z_i is N_i for $i = 1, 2, \ldots, 5$. We assume a total average power constraint P on $X = (X_1, X_2)$.

Using Gaussian signalling it can be easily shown that the BZT region is the set of all (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \mathcal{C}\left(\frac{\alpha P_1}{\bar{\alpha}P_1 + N_1 + N_2 + N_3}\right) + \mathcal{C}\left(\frac{\beta(P - P_1)}{\bar{\beta}(P - P_1) + N_4 + N_5}\right), \\ R_0 &\le \mathcal{C}\left(\frac{\alpha P_1}{\bar{\alpha}P_1 + N_1}\right), \\ R_1 &\le \mathcal{C}\left(\frac{\bar{\alpha}P_1}{N_1 + N_2}\right) + \mathcal{C}\left(\frac{\bar{\beta}(P - P_1)}{N_4}\right), \end{aligned} \tag{12}$$

for some $0 \le P_1 \le P$, $0 \le \alpha, \beta \le 1$.

Now if we use Gaussian signaling to evaluate region (9), one obtains the achievable rate region

consisting of the set of all (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \mathcal{C}\left(\frac{a_1P_1}{\bar{a}_1P_1+N_1+N_2+N_3}\right)+\mathcal{C}\left(\frac{a_2(P-P_1)}{\bar{a}_2(P-P_1)+N_4+N_5}\right), \\ R_0 &\le \mathcal{C}\left(\frac{(a_1+b_1)P_1}{(1-a_1-b_1)P_1+N_1}\right), \\ R_0+R_1 &\le \mathcal{C}\left(\frac{\bar{a}_1P_1}{N_1+N_2}\right)+\mathcal{C}\left(\frac{\bar{a}_2(P-P_1)}{N_4}\right)+\mathcal{C}\left(\frac{a_1P_1}{\bar{a}_1P_1+N_1+N_2+N_3}\right) \\ &\quad +\mathcal{C}\left(\frac{a_2(P-P_1)}{\bar{a}_2(P-P_1)+N_4+N_5}\right), \\ R_0+R_1 &\le \mathcal{C}\left(\frac{((1-a_1-b_1)P_1}{N_1+N_2}\right)+\mathcal{C}\left(\frac{(1-a_2-b_2)(P-P_1)}{N_4}\right) \\ &\quad +\mathcal{C}\left(\frac{(a_1+b_1)P_1}{(1-a_1-b_1)P_1+N_1}\right), \end{aligned} \tag{13}$$

for some $0 \le P_1 \le P$, $0 \le a_1, a_2, b_1, b_2, a_1+b_1, a_2+b_2 \le 1$.
Now consider the above regions with the parameters values: $P = 1, N_1 = 0.4, N_2 = N_3 = 0.1, N_4 = 0.5, N_5 = 0.1$. Fixing $R_1 = 0.5\log(0.49/0.3)$, we can show that the maximum achievable R_0 in the Gaussian BZT region is $0.5\log(2.0566...)$. On the other hand, using the parameter values $b_1 = 0.05, 1-a_1 = 0.1725$, $1-a_2 = 0.5079$, and $P_1 = 0.5680$ for the region given by (13), the pair $(0.5\log(2.0603), 0.5\log(0.49/0.3))$ is in the exterior of the region. Thus restricted to Gaussian signalling the BZT region (8) is strictly contained in region (9). However, we have not been able to prove that Gaussian signaling is optimal for either the BZT region or the capacity region.

2) The reader may ask why we did not consider the more general product channel

$$\begin{aligned} X_1 &\to Y_{21} \to Y_{11} \to Y_{31}, \\ X_2 &\to Y_{12} \to Y_{32} \to Y_{22}. \end{aligned}$$

In fact we considered this more general class at first but were unable to show that the capacity region conditions reduce to the separated form

$$\begin{aligned} R_0 &\le I(V_{11};Y_{31}) + I(V_{12};Y_{32}), \\ R_0 &\le I(V_{21};Y_{21}) + I(V_{22};Y_{22}), \\ R_0+R_1 &\le I(V_{11};Y_{31}) + I(V_{12};Y_{32}) + I(X_1;Y_{11}|V_{11}) + I(X_2;Y_{12}|V_{12}), \\ R_0+R_1 &\le I(V_{21};Y_{21}) + I(V_{22};Y_{22}) + I(X_1;Y_{11}|V_{21}) + I(X_2;Y_{12}|V_{12}), \end{aligned}$$

for some $p(v_{11})p(v_{21}|v_{11})p(x_1|v_{21})p(v_{12})p(v_{22}|v_{12})p(x_2|v_{22})$.

V. Bounds on Capacity of General 3-Receiver Broadcast Channel with Degraded Message Sets

In this section we extend the results in Section III to obtain inner and outer bounds on the capacity region of general 3-receiver broadcast channel with degraded message sets. We first consider the same 2 degraded message set scenario as in Section III but without the condition that $X \to Y_1 \to Y_3$ form a degraded broadcast channel. We establish inner and outer bounds for this case and show that they are tight when the channel $X \to Y_1$ is *less noisy* than the channel $X \to Y_3$, which is a more general class than degraded broadcast channels [13]. We then extend our results to the case of 3 degraded message sets, where M_0 is to be sent to all receivers, M_1 is to be sent to receivers Y_1 and Y_2 and M_2 is to be sent only to receiver Y_1. A special case of this inner bound gives an inner bound to the capacity of the 2 degraded message set scenario where M_0 is to be sent to all receivers and M_1 is to be sent to receivers Y_1 and Y_2 only.

A. *Inner and Outer Bounds for 2 Degraded Message Sets*

We use superposition coding, indirect decoding, and the Marton achievability scheme for the general 2-receiver broadcast channels [14] to establish the following inner bound.

Proposition 5: A rate pair (R_1, R_2) is achievable in a general 3-receiver broadcast channel with 2 degraded message sets if it satisfies the following inequalities:

$$\begin{aligned} R_0 &\leq \min\{I(U_2;Y_2), I(U_3;Y_3)\}, \\ 2R_0 &\leq I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1), \\ R_1 &\leq \min\{I(X;Y_1|U_2) + I(X;Y_1|U_3), I(X;Y_1|U_1)\}, \\ R_0 + R_1 &\leq \min\{I(X;Y_1), I(U_2;Y_2) + I(X;Y_1|U_2), I(U_3;Y_3) + I(X;Y_1|U_3)\}, \\ 2R_0 + R_1 &\leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(U_2;U_3|U_1), \\ 2R_0 + 2R_1 &\leq I(U_2;Y_2) + I(X;Y_1|U_2) + I(U_3;Y_3) + I(X;Y_1|U_3) - I(U_2;U_3|U_1), \end{aligned} \tag{14}$$

for some $p(u_1,u_2,u_3,x) = p(u_1)p(u_2|u_1)p(x,u_3|u_2) = p(u_1)p(u_3|u_1)p(x,u_2|u_3)$ (or in other words, both $U_1 \to U_2 \to (U_3, X)$ and $U_1 \to U_3 \to (U_2, X)$ form Markov chains).

Proof: The general idea of the proof is to represent M_0 by U_1, superimpose two independent pieces of information about M_1 to obtain U_2 and U_3, respectively, and then superimpose the remaining information about M_1 to obtain X. Receiver Y_1 decodes U_1, U_2, U_3, X, receivers Y_2 and Y_3 find M_0 via indirect decoding of U_2 and U_3, respectively, as in Theorem 1. We now provide an outline of the proof

Code Generation: Let $R_1 = S_1 + S_2 + S_3$, where the $S_i \geq 0$, $i = 1,2,3$ and $T_2 \geq S_2$, $T_3 \geq S_3$. Fix a probability mass function of the required form, $p(u_1,u_2,u_3,x) = p(u_1)p(u_2|u_1)p(x,u_3|u_2) = p(u_1)p(u_3|u_1)p(x,u_2|u_3)$.

Generate 2^{nR_0} sequences $U_1^n(m_0)$, $m_0 \in [1, 2^{nR_0}]$ distributed uniformly at random over the set of typical U_1^n sequences. For each $U_1^n(m_0)$ generate 2^{nT_2} sequences $U_2^n(m_0, t_2)$, $t_2 \in [1, 2^{nT_2}]$ distributed uniformly at random from the set of conditionally typical U_2^n sequences, and 2^{nT_3} sequences $U_3^n(m_0, t_3)$, $t_3 \in [1, 2^{nT_3}]$ distributed uniformly at random over the set of conditionally typical U_3^n sequences. Randomly partition the 2^{nT_2} sequences $U_2^n(m_0, t_2)$ into 2^{nS_2} equal size bins and the 2^{nT_3} $U_3^n(m_0, t_3)$ sequences into 2^{nS_3} equal size bins. To ensure that each product bin contains a jointly typical pair $(U_2^n(m_0,t_2), U_3^n(m_0,t_3))$ with arbitrarily high probability, we require that (see [15] for the proof)

$$S_2 + S_3 < T_2 + T_3 - I(U_2;U_3|U_1). \tag{15}$$

Finally for each chosen jointly typical pair $(U_2^n(m_0,t_2), U_3^n(m_0,t_3))$ in each product bin (s_2, s_3), generate 2^{nS_1} sequences of codewords $X^n(m_0, s_2, s_3, s_1)$, $s_1 \in [1, 2^{nS_1}]$ distributed uniformly at random over the set of conditionally typical X^n sequences.

Encoding:

To send the message pair (m_0, m_1), we express m_1 by the triple (s_1, s_2, s_3) and send the codeword $X^n(m_0, s_2, s_3, s_1)$.

Decoding:

1) Receiver Y_1 declares that (m_0, s_2, s_3, s_1) is sent if it is the unique rate tuple such that Y_1^n is jointly typical with $((U_1^n(m_0), U_2^n(m_0,t_2), U_3^n(m_0,t_3), X^n(m_0,s_2,s_3,s_1))$, and s_2 is the product bin number of $U_2^n(m_0,t_2)$ and s_3 is the product bin number of $U_2^n(m_0,t_3)$. Assuming $(m_0, s_1, s_2, s_3) = (1,1,1,1)$ is sent, we partition the error event into the following events.
   a) Error event corresponding to $m_0 \neq 1$ occurs with arbitrarily small probability provided
   $$R_0 + S_1 + S_2 + S_3 < I(X;Y_1). \tag{16}$$
   b) Error event corresponding to $m_0 = 1, s_2 \neq 1, s_3 \neq 1$ occurs with arbitrarily small probability provided
   $$S_1 + S_2 + S_3 < I(X;Y_1|U_1). \tag{17}$$

c) Error event corresponding to $m_0 = 1, s_2 = 1, s_3 \neq 1$ occurs with arbitrarily small probability provided

$$S_1 + S_3 < I(X; Y_1|U_1, U_2) = I(X; Y_1|U_2). \tag{18}$$

The equality follows from the fact that $U_1 \to U_2 \to (U_3, X)$ form a Markov Chain.

d) Error event corresponding to $m_0 = 1, s_2 \neq 1, s_3 = 1$ occurs with arbitrarily small probability provided

$$S_1 + S_2 < I(X; Y_1|U_1, U_3) = I(X; Y_1|U_3). \tag{19}$$

The above equality uses the fact that $U_1 \to U_3 \to (U_2, X)$ forms a Markov chain.

e) Error event corresponding to $m_0 = 1, s_2 = 1, s_3 = 1, s_1 \neq 1$ occurs with arbitrarily small probability provided

$$S_1 < I(X; Y_1|U_1, U_2, U_3) = I(X; Y_1|U_2, U_3). \tag{20}$$

Note that the equality here uses a weaker Markov structure $U_1 \to (U_2, U_3) \to X$.

Thus receiver Y_1 decodes (m_0, s_2, s_3, s_1) with arbitrarily small probability of error provided equations (16)-(20) hold.

2) Receiver Y_2 decodes m_0 via list decoding of $U_2^n(m_0, t_2)$ (as in Theorem 1). This can be achieved with arbitrarily small probability of error provided

$$R_0 + T_2 < I(U_2; Y_2). \tag{21}$$

3) Receiver Y_3 decodes m_0 via list decoding of $U_2^n(m_0, t_3)$ (as in Theorem 1). This can be achieved with arbitrarily small probability of error provided

$$R_0 + T_3 < I(U_3; Y_3). \tag{22}$$

Combining equations (15)-(22) and using the Fourier-Motzkin procedure [16] to eliminate T_2, T_3, S_1, S_2, and S_3, we obtain the inequalities in (14). The details are given in Appendix II. ■

Remarks:

1) The above achievability scheme can be adapted to any joint distribution $p(u_1, u_2, u_3, x)$. However by letting $\tilde{U}_2 = (U_2, U_1)$ and letting $\tilde{U}_3 = (U_3, U_1)$ we observe that the region remains unchanged. Hence, without loss of generality we assume the structure of the auxiliary random variables as described in the proposition. It is also interesting to note that the auxiliary random variables in the outer bound described in the next subsection also possess the same structure.
2) An interesting choice of the auxiliary random variables is to set U_2 or U_3 equal to U_1 (i.e., only one of the the receivers tries to indirectly decode M_0), say let $U_3 = U_1$. This reduces the inequalities 5 (after removing the redundant ones) to:

$$\begin{aligned} R_0 &\leq \min\{I(U_2; Y_2), I(U_1; Y_3)\}, \\ R_1 &\leq I(X; Y_1|U_1), \\ R_0 + R_1 &\leq \min\{I(X; Y_1), I(U_2; Y_2) + I(X; Y_1|U_2), I(U_1; Y_3) + I(X; Y_1|U_1)\}, \end{aligned} \tag{23}$$

where $U_1 \to U_2 \to X$ form a Markov chain.
This region includes the capacity region of the multilevel case in Theorem 1. It is easy to verify that for any $U_1 \to U_2 \to X$ that form a Markov chain, the corner points of the region in Theorem 1 satisfy the above inequalities (and this suffices by convexity).

We now establish the following outer bound

Proposition 6: Any achievable rate pair (R_0, R_1) for the general 3-receiver broadcast channel with 2 degraded message sets must satisfy the conditions:

$$
\begin{aligned}
R_0 &\leq \min\{I(U_1;Y_1), I(U_2;Y_2) - I(U_2;Y_1|U_1), I(U_3;Y_3) - I(U_3;Y_1|U_1)\},\\
R_1 &\leq I(X;Y_1|U_1).
\end{aligned}
$$

for some $p(u_1,u_2,u_3,x) = p(u_1)p(u_2|u_1)p(x,u_3|u_2) = p(u_1)p(u_3|u_1)p(x,u_2|u_3)$, i.e., the same structure of the auxiliary random variables as in Lemma 5. Further one can restrict the cardinalities of U_1, U_2, U_3 to: $\|\mathcal{U}_1\| \leq \|\mathcal{X}\| + 6$, $\|\mathcal{U}_2\| \leq (\|\mathcal{X}\|+1)(\|\mathcal{X}\|+6)$, and $\|\mathcal{U}_3\| \leq (\|\mathcal{X}\|+1)(\|\mathcal{X}\|+6)$.

Proof: The proof follows largely standard arguments. The auxiliary random variables are identified as $U_{1i} = (M_0, Y_1^{i-1})$, $U_{2i} = (U_{1i}, Y_2{}_{i+1}^{n})$, $U_{3i} = (U_{1i}, Y_3{}_{i+1}^{n})$. With this identification inequalities $R_0 \leq I(U_1;Y_1)$ and $R_1 \leq I(X;Y_1|U_1)$ is immediate. The other two inequalities also follow from standard arguments and is briefly outlined here.

$$
\begin{aligned}
nR_0 &\leq n\lambda_n + \sum_i I(M_0; Y_{3i}|Y_3{}_{i+1}^{n})\\
&\leq n\lambda_n + \sum_i I(M_0, Y_3{}_{i+1}^{n}, Y_1^{i-1}; Y_{3i}) - I(Y_1^{i-1}; Y_{3i}|M_0, Y_3{}_{i+1}^{n})\\
&\overset{(a)}{=} n\lambda_n + \sum_i I(M_0, Y_3{}_{i+1}^{n}, Y_1^{i-1}; Y_{3i}) - I(Y_3{}_{i+1}^{n}; Y_{1i}|M_0, Y_1^{i-1})\\
&= n\lambda_n + \sum_i I(U_{3i};Y_{3i}) - I(U_{3i};Y_{1i}|U_{1i}),
\end{aligned}
$$

where (a) uses the Csiszár sum equality.

The cardinality bounds are established using a similar argument as in III-C. To create a set of new auxiliary random variables with the bounds of Proposition 6, we first replace U_2 by (U_2, U_1) and U_3 by (U_3, U_1). It is easy to see from the Markov chain relationships $U_1 \to U_2 \to (U_3, X)$ and $U_1 \to U_3 \to (U_2, X)$ that the following region is same as the that of Proposition 6.

$$
\begin{aligned}
R_0 \leq \min\{&I(U_1;Y_1), I(U_1,U_2;Y_2) + I(X;Y_1|U_1,U_2) - I(X:Y_1|U_1),\\
&I(U_1,U_3;Y_3) + I(X;Y_1|U_1,U_3) - I(X:Y_1|U_1)\},\\
R_1 \leq I&(X;Y_1|U_1).
\end{aligned} \tag{24}
$$

Then using standard arguments one can replace U_1 by U_1^* satisfying $\|\mathcal{U}_1^*\| \leq \|\mathcal{X}\| + 6$, such that the distribution of X and $H(Y_1|U_1)$, $H(Y_1|U_1,U_2)$, $H(Y_1|U_1,U_3)$, $H(Y_2|U_1)$, $H(Y_2|U_1,U_2)$, $H(Y_3|U_1)$, and $H(Y_3|U_1,U_3)$ are preserved. Now for each $U_1^* = u_1$ one can find $U_2^*(u_1)$ with cardinality less than $\|\mathcal{X}\|+1$ each such that the distribution of X conditioned on $U_1^* = u_1$, $H(Y_1|U_1^* = u_1, U_2)$, and $H(Y_2|U_1^* = u_1, U_2)$ are preserved. Similarly one can find for each $U_1^* = u_1$, a random variable $U_3^*(u_1)$ with cardinality less than $\|\mathcal{X}\|+1$ each such that the distribution of X conditioned on $U_1^* = u_1$, $H(Y_1|U_1^* = u_1, U_3)$, and $H(Y_3|U_1^* = u_1, U_3)$ are preserved. This yields random variables U_1^*, U_2^*, U_3^* that preserve the region in (24). (Note that as the distribution of X conditioned on $U_1 = u_1$ is preserved by both $U_2^*(u_1)$ and $U_3^*(u_1)$, it is possible to get a consistent triple of random variables U_1^*, U_2^*, U_3^*.) Finally setting $\tilde{U}_1 = U_1^*$, $\tilde{U}_2 = (U_1^*, U_2^*)$ and $\tilde{U}_3 = (U_1^*, U_3^*)$ gives the desired bounds on cardinality as well as the desired Markov relations. ■

Remarks:

1) The above outer bound appears to be very different from the inner bound of Proposition 5. However, by taking appropriate sums of the inequalities defining the region of Proposition 6, we arrive at the

conditions

$$\begin{aligned}
R_0 &\le \min\{I(U_2;Y_2) - I(U_2;Y_1|U_1), I(U_3;Y_3) - I(U_3;Y_1|U_1)),\\
R_1 &\le I(X;Y_1|U_1),\\
R_0 + R_1 &\le \min\{I(X;Y_1), I(U_2;Y_2) + I(X;Y_1|U_2), I(U_3;Y_3) + I(X;Y_1|U_3)\},\\
2R_0 + R_1 &\le I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(U_2;U_3|U_1) + I(U_2;U_3|Y_1,U_1).
\end{aligned}$$

These conditions include some redundant ones, but are closer in structure to the inequalities defining the inner bound of Proposition 5.

2) The outer bound in Proposition 6 reduces to the capacity region for the multilevel case in Theorem 1. To see this observe that when $X \to Y_1 \to Y_3$ form a Markov chain,

$$R_0 \le I(U_3;Y_3) - I(U_3;Y_1|U_1) \le I(U_3;Y_3) - I(U_3;Y_3|U_1) = I(U_1;Y_3). \tag{25}$$

Further from $R_1 \le I(X;Y_1|U_1)$, we have $R_0 + R_1 \le I(U_1;Y_3) + I(X;Y_1|U_1)$. Thus the outer bound is contained in the achievable region of Theorem 1, i.e.,

$$\begin{aligned}
R_0 &\le \min\{I(U_1;Y_3), I(U_2;Y_2)\}, \qquad (26)\\
R_0 + R_1 &\le \{I(U_1;Y_3) + I(X;Y_1|U_1), I(U_2;Y_2) + I(X;Y_1|U_2)\}.
\end{aligned}$$

3) The inner and outer bounds match if Y_1 is less noisy than Y_3 [13], that is if $I(U;Y_3) \le I(U;Y_1)$ for all $p(u)p(x|u)$. As shown in [13], this condition is more general than degradedness. As such, it defines a larger class than multilevel broadcast channels.

Proposition 7: The capacity region for the 3-receiver broadcast channel with 2 degraded message sets when Y_1 is a *less noisy* receiver than Y_3 is given by the set of rate pairs (R_0, R_1) such that

$$\begin{aligned}
R_0 &\le \min\{I(U_1;Y_3), I(U_2;Y_2)\}, \qquad (27)\\
R_0 + R_1 &\le \min\{I(U_1;Y_3) + I(X;Y_1|U_1), I(U_2;Y_2) + I(X;Y_1|U_2)\},
\end{aligned}$$

for some $p(u_1)p(u_2|u_1)p(x|u_2)$.

From the definition of less noisy receivers [13] we have $I(U_3;Y_3|U_1 = u_1) \le I(U_3;Y_1|U_1 = u_1)$ for every choice of u_1 and thus $I(U_3;Y_3|U_1) \le I(U_3;Y_1|U_1)$ for every $p(u_1)p(u_3|u_1)p(x|u_3)$. Using (25) it follows that the general outer bound is contained in (27). The corner point of (27) (under the less noisy assumption) is contained in the region given by (23) and thus achievable by setting $U_3 = U_1$ in the region of Proposition 5.

B. Inner Bound for 3 Degraded Message Sets

In this section we establish an inner bound to the capacity region of the broadcast channel with 3 degraded message sets where M_0 is to be sent to all three receivers, M_1 is to be sent only to Y_1 and Y_2, and M_2 is to be sent only to Y_1. We then specialize the result to the case of 2 degraded message sets scenario, where M_0 is to be sent to all receivers and M_1 is to be sent to Y_1 and Y_2 and establish optimality for two classes of channels.

The inner bound we establish is closely related to that of Proposition 5. To explain the connection, consider a 3-receiver broadcast channel scenario where message M_0 is to be sent to all three receivers, message M_{12} is to be sent to receivers Y_1 and Y_2, message M_{13} is to be sent to receivers Y_1 and Y_3, and message M_{11} is to be sent only to receiver Y_1. An inner bound to the capacity region for this scenario that uses superposition coding and Marton's coding scheme would be to represent M_0 by an auxiliary random variable U_1, (M_0, M_{12}) by an auxiliary random variable U_2, (M_0, M_{13}) by U_3, and $(M_0, M_{12}.M_{13}, M_{11})$ by X, where $U_1 \to U_2 \to (U_3, X)$ and $U_1 \to U_3 \to (U_2, X)$ form Markov chains.

The inner bound of Proposition 5 follows from the above scenario by relaxing the conditions that Y_2 needs to decode M_{12} and Y_3 needs to decode M_{13} and considering both messages as parts of the

private message to receiver Y_1. However, instead of eliminating the auxiliary random variables U_2 and U_3 completely (as in the BZT region, which is a straightforward extension of the Körner-Marton scheme), we keep them and have receivers Y_2 and Y_3 use the new technique of indirect decoding to find M_0 through U_2 and U_3, respectively. As we have shown in Section IV, having these random variables U_2 and U_3 can strictly improve the achievability region of the 2-message sets scenario.

Now consider the 3 degraded message set scenario. We relax the condition in the above scenario that Y_3 needs to decode M_{13}. Recall the proof of Proposition 5. We let $R_1 = S_2$, $R_2 = S_3 + S_1$ and represent M_0 by U_1, (M_0, M_1) by U_2, (M_0, S_3) by U_3, and (M_0, M_1, M_2) by X. Receiver Y_1 finds (M_0, M_1, M_2) by decoding U_1, U_2, U_3, X; receiver Y_2 finds (M_0, M_1) by decoding U_1, U_2; and receiver Y_3 finds M_0 by indirectly decoding U_1 through U_3. We obtain the following conditions for achievability of any rate tuple (R_0, R_1, S_3, S_1) by replacing S_2 by R_1 in conditions (15)-(22) and adding the condition $T_2 < I(U_2; Y_2|U_1)$ (to enable Y_2 to completely decode U_2).

$$\begin{aligned}
R_1 &\le T_2,\\
S_3 &\le T_3,\\
R_1 + S_3 &\le T_2 + T_3 - I(U_2; U_3|U_1),\\
R_0 + S_1 + R_1 + S_3 &\le I(X; Y_1),\\
S_1 + S_3 &\le I(X; Y_1|U_1, U_2) = I(X; Y_1|U_2),\\
S_1 + R_1 &\le I(X; Y_1|U_1, U_3) = I(X; Y_1|U_3),\\
S_1 + R_1 + S_3 &\le I(X; Y_1|U_1),\\
S_1 &\le I(X; Y_1|U_1, U_2, U_3) = I(X; Y_1|U_2, U_3),\\
R_0 + T_2 &\le I(U_1, U_2; Y_2) = I(U_2; Y_2),\\
T_2 &\le I(U_2; Y_2|U_1),\\
R_0 + T_3 &\le I(U_1, U_3; Y_3) = I(U_3; Y_3),
\end{aligned} \tag{28}$$

for some $p(u_1, u_2, u_3, x) = p(u_1)p(u_2|u_1)p(x, u_3|u_2) = p(u_1)p(u_3|u_1)p(x, u_2|u_3)$.

Performing Fourier-Motzkin procedure to eliminate the variables S_1, S_3, T_2 and T_3 yields the following achievable region.

Theorem 2: A rate triple (R_0, R_1, R_2) is achievable in a general 3-receiver broadcast channel with 3 degraded message sets if it satisfies the conditions:

$$\begin{aligned}
R_0 &\le I(U_3; Y_3)\\
R_1 &\le \min\{I(U_2; Y_2|U_1), I(X; Y_1|U_3)\},\\
R_2 &\le I(X; Y_1|U_2)\\
R_0 + R_1 &\le \min\{I(U_2; Y_2), I(U_2; Y_2|U_1) + I(U_3; Y_3) - I(U_2; U_3|U_1)\},\\
2R_0 + R_1 &\le I(U_2; Y_2) + I(U_3; Y_3) - I(U_2; U_3|U_1),\\
R_0 + R_2 &\le I(U_3; Y_3) + I(X; Y_1|U_2, U_3)\\
R_1 + R_2 &\le I(X; Y_1|U_1),\\
R_0 + R_1 + R_2 &\le \min\{I(X; Y_1), I(U_3; Y_3) + I(X; Y_1|U_3),\\
&\qquad I(U_2; Y_2|U_1) + I(U_3; Y_3) + I(X; Y_1|U_2, U_3) - I(U_2; U_3|U_1)\},\\
2R_0 + R_1 + R_2 &\le I(U_2; Y_2) + I(U_3; Y_3) + I(X; Y_1|U_2, U_3) - I(U_2; U_3|U_1),\\
R_0 + 2R_1 + R_2 &\le I(U_2; Y_2|U_1) + I(U_3; Y_3) + I(X; Y_1|U_3) - I(U_2; U_3|U_1),\\
2R_0 + 2R_1 + R_2 &\le I(U_2; Y_2) + I(U_3; Y_3) + I(X; Y_1|U_3) - I(U_2; U_3|U_1).
\end{aligned} \tag{29}$$

for some $p(u_1, u_2, u_3, x) = p(u_1)p(u_2|u_1)p(x, u_3|u_2) = p(u_1)p(u_3|u_1)p(x, u_2|u_3)$ (i.e., as before both $U_1 \to U_2 \to (U_3, X)$ and $U_1 \to U_3 \to (U_2, X)$ form Markov chains).

Remark: The region of Theorem 2 reduces to the inner bound of Proposition 5 by setting $R_1 = 0$. The equivalence between the two descriptions is proved in Appendix III.

We now consider a 2 degraded message set scenario where M_0 is to be sent to all receivers and M_1 is to be sent to receivers Y_1 and Y_2. The following inner bound follows from Theorem 2 by setting $R_2 = 0$.

Corollary 1: A rate pair (R_0, R_1) is achievable in a 3-receiver broadcast channel with 2 degraded message sets, where M_0 is to be decoded by all three receivers and M_1 is to be decoded only by Y_1 and Y_2 if it satisfies the following conditions:

$$\begin{aligned} R_0 &\leq I(U;Y_3), \\ R_1 &\leq \min\{I(X;Y_2|U), I(X;Y_1|U)\}, \\ R_0 + R_1 &\leq \min\{I(X;Y_2), I(X;Y_1)\}, \end{aligned} \tag{30}$$

for some $p(u)p(x|u)$.

Remarks:

1) Region (30) coincides with the straightforward extension of the Körner-Marton 2-receiver region.
2) By setting $R_2 = 0$, $U_2 = X$, and $U_3 = U_1 = U$ the region in Theorem 2 reduces to (30). Thus region in (30) is contained in region (29).
3) It may seem that the region obtained by setting $R_2 = 0$ in (29) is larger than region (30), but they are in fact equal. variables, we see that Therefore, there is no need to introduce U_3. To prove this, observe that

$$\begin{aligned} R_0 + R_1 &\leq I(U_2;Y_2|U_1) + I(U_3;Y_3) - I(U_2;U_3|U_1) \\ &= I(U_3;Y_3) + I(U_3;Y_2|U_1) + I(U_2;Y_2|U_3) - I(U_3;Y_2|U_2) - I(U_3;U_2|U_1) \\ &= I(U_3;Y_3) + I(U_2;Y_2|U_3) - I(U_3;U_2|Y_2,U_1) \\ &\leq I(U_3;Y_3) + I(X;Y_2|U_3). \end{aligned}$$

Thus the rate pairs must satisfy the following inequalities

$$\begin{aligned} R_0 &\leq I(U_3;Y_3), \\ R_0 + R_1 &\leq \min\{I(U_3;Y_3) + I(X;Y_2|U_3), I(U_3;Y_3) + I(X;Y_1|U_3)\}, \\ R_0 + R_1 &\leq \min\{I(X;Y_2), I(X;Y_1)\}. \end{aligned} \tag{31}$$

Clearly this is contained inside region (30) and hence region (29) reduces to the one in Corollary 1 when $R_2 = 0$.
4) Inner bound (30) is optimal for the following two special classes of broadcast channels.

Proposition 8: Achievable region (30) is tight for deterministic 3-receiver broadcast channels.

It is straightforward to show that the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{H(Y_1), H(Y_2), H(Y_3)\}, \\ R_0 + R_1 &\leq \min\{H(Y_2), H(Y_1)\}, \end{aligned}$$

for some $p(x)$ constitutes an outer bound on the capacity region. To show achievability, we need only consider the three choices for U: (i) $U = Y_3$, and (ii) $U = X$, and (iii) $U = \emptyset$.

Proposition 9: Achievable region (30) is optimal when Y_1 is a less noisy receiver than Y_3 and Y_2 is a less noisy receiver than Y_3.

Note that this result generalizes Theroem 3.2 in [4] where the authors assume the receivers are Y_2 and Y_1 are degraded versions of Y_3. To show optimality, we set $U_i = (M_0, Y_3^{i-1})$ and thus the only

non-trivial inequality in the converse is $R_1 \leq \min\{I(X;Y_1|U), I(X;Y_2|U)\}$. To see this observe that

$$\begin{aligned}
nR_1 &\leq \sum_i I(M_1; Y_{1i}|M_0, Y_1{}^n{}_{i+1}) \\
&\leq \sum_i I(M_1; Y_{1i}|M_0, Y_1{}^n{}_{i+1}, Y_3^{i-1}) + \sum_i I(Y_3^{i-1}; Y_{1i}|M_0, Y_1{}^n{}_{i+1})) \\
&\overset{(a)}{=} \sum_i I(M_1, Y_1{}^n{}_{i+1}; Y_{1i}|M_0, Y_3^{i-1}) - \sum_i I(Y_1{}^n{}_{i+1}; Y_{1i}|M_0, Y_3^{i-1}) + \sum_i I(Y_1{}^n{}_{i+1}; Y_{3i}|M_0, Y_3^{i-1}) \\
&\overset{(b)}{\leq} \sum_i I(X_i; Y_{1i}|M_0, Y_3^{i-1}),
\end{aligned}$$

where (a) uses the Csiszár sum equality and (b) uses the assumption that Y_1 is a less noisy than Y_3, which implies that $I(Y_1{}^n{}_{i+1}; Y_{3i}|M_0, Y_3^{i-1}) \leq I(Y_1{}^n{}_{i+1}; Y_{1i}|M_0, Y_3^{i-1})$. The bound $R_1 \leq I(X;Y_2|U)$ can be proved similarly.

C. Inner Bounds for k-receiver Broadcast Channels

The inner bounds discussed in previous subsections suggest the following extension to general k-receiver broadcast channel scenarios with given message requirements. To illustrate our procedure we shall use the running example of a 3-receiver broadcast channel with 3 messages to receiver subsets: $\{1\}$, $\{1,2\}$, and $\{2,3\}$.

To obtain an inner bound to capacity for a given message requirement, we first consider all nonempty receiver subsets. Let $\mathcal{S}_D$ be the collection of subsets specified by the message requirements. For each $A \in \mathcal{S}_D$, we introduce an auxiliary random variable for every $B \supset A$. Thus in our example, $\mathcal{S}_D = \{\{1\},\{2,3\},\{1,2\}\}$, and five auxiliary random variables are introduced corresponding to the subsets: $\{1,2,3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, and $\{1\}$. Let $\mathcal{S}_I$ denote the receiver subsets for which auxiliary random variables are introduced but are not in $\mathcal{S}_D$. In the example, $\mathcal{S}_I = \{\{1,3\},\{1,2,3\}\}$.

The receiver subsets with auxiliary random variables assigned to them are classified into levels based on their cardinality with the lowest level subsets having the largest cardinality. There is a Markov structure between the variables as follows: if U_B represents the auxiliary random variable corresponding to the subset B and U_A represents the auxiliary random variable corresponding to the subset $A \subset B$, then one can set $U_A = (U_B, \tilde{U}_A)$. Thus an auxiliary random variable U_A corresponding to a subset A should contain *all* auxiliary random variables corresponding to the subsets $B \supset A$. For the running example, Level 1 contains the subset $\{1,2,3\}$, Level 2 contains the subsets $\{1,2\}$, $\{1,3\}$, and $\{2,3\}$, and Level 3 contains the subset $\{1\}$. The Markov relationships between these auxiliary random variables are defined by:

$$\begin{aligned}
&U_{12} = (U_{123}, \tilde{U}_{12}),\ U_{13} = (U_{123}, \tilde{U}_{13}),\ U_{23} = (U_{123}, \tilde{U}_{23}), \\
&U_1 = \{U_{12}, U_{13}, \tilde{U}_1\}.
\end{aligned}$$

Code generation proceeds one level at a time beginning with the lowest level followed by the second lowest level, and so on. The codebooks corresponding to auxiliary random variables at each level are randomly generated conditioned on codewords at the lower level according to the Markov structure of the auxiliary random variables. Random binning is performed at each level to find jointly typical codewords to represent message products.

Decoding is performed at receiver i as follows: let T_i represent the collection of receiver subsets that contains i for which auxiliary random variables are introduced. A subset $A \in T_i$ is said to be minimal if there is no $B \in T_i$ such that $B \subset A$. Let $\mathcal{T}_i^{\min}$ be the collection of minimal subsets in T_i. For the example

we obtain

$$T_1 = \{\{1,2,3\},\{1,2\},\{1,3\},\{1\}\} \text{ and } \mathcal{T}_1^{\min} = \{1\},$$
$$T_2 = \{\{1,2,3\},\{1,2\},\{2,3\}\} \text{ and } \mathcal{T}_2^{\min} = \{\{1,2\},\{2,3\}\},$$
$$T_3 = \{1,2,3\},\{1,3\},\{2,3\}\} \text{ and } \mathcal{T}_3^{\min} = \{\{1,3\},\{2,3\}\}.$$

By the Markov structure defined above, it is clear that all the messages for receiver i are represented by the auxiliary random variables $\mathcal{U}_i^{\min}$, which correspond to the elements of $\mathcal{T}_i^{min}$. The auxiliary random variables in $\mathcal{U}_i^{min} \cap \mathcal{S}_D$ represent private messages for receiver i, while those in $\mathcal{U}_i^{min} \cap \mathcal{S}_D^c$ contain only parts of private messages. Receiver i uses indirect decoding to find the private messages encoded into cloud centers by using the satellite codewords represented by the auxiliary random variables in $\mathcal{U}_i^{min}$.

In our running example, receiver 3 indirectly decodes U_{23} using the pair (U_{13}, U_{23}). That is, the rate constraints are such that receiver 3 may not be able to uniquely decode U_{13} but is able to decode the correct U_{23}. However, receivers Y_2 and Y_1 should be able to correctly decode (U_{12}, U_{23}) and U_1, respectively, and hence these receivers impose the usual (direct) decoding constraints on the rates. In general, when $\mathcal{U}_i^{min} \cap \mathcal{S}_I = \emptyset$, indirect decoding is not needed as in the Examples below, where as in Proposition 5 indirect decoding is needed.

The following two examples show that the above procedure yields the best known inner bounds for special classes of broadcast channels.

Example 1: 2-receiver broadcast channel where M_1 is to be decoded by receiver Y_1 and M_2 is to be decoded by Y_2. We generate 3 auxiliary random variables corresponding to the three non-empty subsets of $\{1,2\}$: W for $\{1,2\}$, U for $\{1\}$ and V for $\{2\}$. Setting $\tilde{U} = (U, W)$ and $\tilde{V} = (V, W)$ represents the Markov structure among the variables. Observe that the auxiliary random variables are exactly as in Marton's coding scheme and so is the code generation we outlined earlier.

Example 2: k-receiver broadcast channel with 2 degraded message sets, where M_0 is to be decoded by receivers $\{1,\dots,k\}$ and M_1 is to be decoded by $\{1,\dots,k-1\}$. The only subsets that we would assign auxiliary random variables to here are $\{1,\dots,k\}$ and $\{1,\dots,k-1\}$. We thus introduce the auxiliary random variable U_1 for $\{1,\dots,k\}$ and U_2 for $\{1,\dots,k-1\}$. The region is then be given by

$$\begin{aligned} R_0 &\le I(U_1; Y_k), \\ R_0 + R_1 &\le I(U_2; Y_i), \text{ for } i = 1,\dots,k-1, \\ R_1 &\le I(U_2; Y_i | U_1) \text{ for } i = 1,\dots,k-1, \end{aligned}$$

where $U_1 \to U_2 \to X \to \{Y_1,\dots,Y_l\}$ form a Markov chain. Clearly in this case it is optimal to set $U_2 = X$, which reduces the region to the straightforward extension of the Körner-Marton scheme.

Remark: Our procedure can result in an explosion in the number of auxiliary random variables introduced even in simple scenarios. However, as we have shown in Section IV, indirect decoding may be needed to achieve the capacity region for some classes of channels. Thus the introduction of such a large number of auxiliary random variables may indeed be necessary in general.

VI. Conclusion

Recent results and conjectures on the capacity of $(k > 2)$-receiver broadcast channels with degraded message sets [6], [4], [5] have lent support to the general belief that the straightforward extension of the Körner-Marton region for the 2-receiver case is optimal. This paper shows that this is not the case. We show that the capacity region of the 3-receiver broadcast channels with 2 degraded message sets can be strictly larger than the straightforward extension of the Körner-Marton region. The achievability proof uses the new idea of indirect decoding whereby a receiver decodes a cloud center indirectly through joint typicality with a satellite codeword. Using this idea, we devise new inner bounds to the capacity of the

general 3-receiver broadcast channel with 2 and 3 degraded message sets and show optimality in some cases. The structure of the auxiliary random variables in the inner bounds can be naturally extended to more than 3 receivers. The bounds also provide some insight into how the Marton achievable rate region may be extended to more than 2 receivers.

The results in this paper suggest that the capacity of the $k > 2$-receiver broadcast channels with degraded message sets is as at least as hard to find as the capacity of the general 2-receiver broadcast channel with common and private message. However, it would be interesting to explore the optimality of our new inner bounds for classes where capacity is known for the general 2-receiver case, such as deterministic and vector Gaussian broadcast channels. It would also be interesting to investigate applications of indirect decoding to other problems, for example, 3-receiver broadcast channels with confidential message sets [11].

ACKNOLEDGEMENT

The authors wish to thank Young-Han Kim for valuable suggestions that has improved the presentation of this paper.

REFERENCES

[1] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Info. Theory*, vol. IT-23, pp. 60–64, Jan, 1977.
[2] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.
[3] J. Körner and K. Marton, "Images of a set via two channels and their role in mult-user communication," *IEEE Trans. Info. Theory*, vol. IT-23, pp. 751–761, Nov, 1977.
[4] S. Diggavi and D. Tse, "On opportunistic codes and broadcast codes with degraded message sets," *Information theory workshop (ITW)*, 2006.
[5] V. Prabhakaran, S. Diggavi, and D. Tse, "Broadcasting with degraded message sets: A deterministic approach," *Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing*, 2007.
[6] S. Borade, L. Zheng, and M. Trott, "Multilevel broadcast networks," *International Symposium on Information Theory*, 2007.
[7] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredac. Inform.*, vol. 10(3), pp. 3–14, 1974.
[8] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 166–169, March, 1979.
[9] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.
[10] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Interscience, 1991.
[11] I. Csizár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Info. Theory*, vol. IT-24, pp. 339–348, May, 1978.
[12] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-21(6), pp. 629–637, November, 1975.
[13] J. Körner and K. Marton, "A source network problem involving the comparison of two channels ii," *Trans. Colloquim Inform. Theory, Keszthely, Hungary*, August, 1975.
[14] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 306–311, May, 1979.
[15] A. El Gamal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 120–121, 1981.
[16] A. Schrijver, *Theory of Integer and Linear Programming*. John Wiley & sons, 1986.

APPENDIX I
PROOF OF PROPOSITIONS 1, 2, 3, AND 4

To prove Propositions 1, 2, note that it is straightforward to show that each simplified characterization is contained in the original region as the characterizations are obtained by using the channels independently. So we only prove the other non-trivial direction.

Proof of Proposition 1:

We prove that for the product broadcast channel given by (7) the BZT region (3) reduces to the expression (8).

Consider the first term (8a) in the BZT region

$$\begin{aligned} R_0 \leq I(U;Y_3) &= I(U;Y_{31},Y_{32}) \\ &= I(U;Y_{31}) + I(U;Y_{32}|Y_{31}) \\ &\leq I(U;Y_{31}) + I(U,Y_{31};Y_{32}) \\ &\leq I(U;Y_{31}) + I(U,Y_{11};Y_{32}). \end{aligned}$$

Now set $V_1 = U$ and $V_2 = (U, Y_{11})$. Thus the above inequality becomes

$$R_0 \leq I(V_1;Y_{31}) + I(V_2;Y_{32}).$$

The second term (8b) in the BZT region is simply given by

$$R_0 \leq I(V_1;Y_{21}).$$

Finally, consider the last term (8c)

$$\begin{aligned} R_1 \leq I(X;Y|U) &= I(X_1,X_2;Y_{11},Y_{12}|U) \\ &= H(Y_{11},Y_{12}|U) - H(Y_{11},Y_{12}|X_1,X_2,U) \\ &= H(Y_{11}|U) + H(Y_{12}|U,Y_{11}) - H(Y_{11}|X_1,U) - H(Y_{12}|X_2,U) \\ &= I(X_1;Y_{11}|U) + H(Y_{12}|U,Y_{11}) - H(Y_{12}|X_2,U,Y_{11}) \\ &= I(X_1;Y_{11}|V_1) + I(X_2;Y_{12}|V_2). \end{aligned}$$

The fact that $p(v_1)p(v_2)p(x_1|v_1)p(x_2|v_2)$ suffices follows from the structure of the mutual information terms.

Proof of Proposition 2:

We prove that for the product broadcast channel (7) the capacity region given by Theorem 1 reduces to the expression (9).

Consider the first term (9a) in the capacity region

$$\begin{aligned} R_0 \leq I(U_1;Y_3) &= I(U_1;Y_{31},Y_{32}) \\ &= I(U_1;Y_{31}) + I(U_1;Y_{32}|Y_{31}) \\ &\leq I(U_1;Y_{31}) + I(U_1,Y_{31};Y_{32}) \\ &\leq I(U_1;Y_{31}) + I(U_1,Y_{11};Y_{32}). \end{aligned}$$

Now set $V_{11} = U_1$ and $V_{12} = (U_1, Y_{11})$.

The second term (9b) in the capacity region is $R_0 \leq I(U_2;Y_{21})$. Now set $V_{21} = U_2$ and from $U_1 \rightarrow U_2 \rightarrow (X_1, X_2)$ we have $V_{11} \rightarrow V_{21} \rightarrow X_1$. Thus the second term can be rewritten as $R_0 \leq I(V_{21};Y_{21})$

Consider the third term (9c)

$$\begin{aligned}R_0 + R_1 &\leq I(U_1; Y_3) + I(X; Y_1|U_1)\\&= I(U_1; Y_{31}, Y_{32}) + I(X_1, X_2; Y_{11}, Y_{12}|U_1)\\&\leq I(U_1; Y_{31}) + I(U_1, Y_{11}; Y_{32}) + H(Y_{11}, Y_{12}|U_1) - H(Y_{11}, Y_{12}|X_1, X_2, U_1)\\&= I(U_1; Y_{31}) + I(U_1, Y_{11}; Y_{32}) + H(Y_{11}|U_1)\\&\qquad + H(Y_{12}|U_1, Y_{11}) - H(Y_{11}|X_1, U_1) - H(Y_{12}|X_2, U_1, Y_{11})\\&= I(V_{11}; Y_{31}) + I(V_{12}; Y_{32}) + I(X_1; Y_{11}|V_{11}) + I(X_2; Y_{12}|V_{12}).\end{aligned}$$

Finally consider the last term (9d)

$$\begin{aligned}R_0 + R_1 &\leq I(U_2; Y_{21}) + I(X; Y_1|U_2)\\&= I(U_2; Y_{21}) + I(X_1, X_2; Y_{11}, Y_{12}|U_2)\\&= I(U_2; Y_{21}) + H(Y_{11}, Y_{12}|U_2) - H(Y_{11}, Y_{12}|X_1, X_2, U_2)\\&\leq I(U_2; Y_{21}) + H(Y_{11}|U_2) + H(Y_{12}|U_2, Y_{11}) - H(Y_{11}|X_1, U_2) - H(Y_{12}|X_2, U_2, Y_{11})\\&= I(V_{21}; Y_{21}) + I(X_1; Y_{11}|V_{21}) + I(X_2; Y_{12}|U_2, Y_{11})\\&= I(V_{21}; Y_{21}) + I(X_1; Y_{11}|V_{21}) + I(X_2; Y_{12}|U_2, U_1, Y_{11})\\&\leq I(V_{21}; Y_{21}) + I(X_1; Y_{11}|V_{21}) + I(U_2, X_2; Y_{12}|U_1, Y_{11})\\&= I(V_{21}; Y_{21}) + I(X_1; Y_{11}|V_{21}) + I(X_2; Y_{12}|V_{12}).\end{aligned}$$

The fact that $p(v_{11})p(v_{21})p(x_1|v_{21})p(v_{12})p(x_2|v_{12})$ suffices follows from the structure of the mutual information terms.

In the proof of propositions 3 and 4 we shall make use of the following simple fact about the entropy function [10].

$$H(ap, 1-p, (1-a)p) = H(p, 1-p) + pH(a, 1-a).$$

Proof of Proposition 3:

We prove that the region given by (8) reduces to (10) for the binary erasure channel described by the example in Section IV.

Let $\mathrm{P}\{V_1 = i\} = \alpha_i,\ \mathrm{P}\{X_1 = 0|V_1 = i\} = \mu_i$. Then,

$$\begin{aligned}I(V_1; Y_{31}) &= H\left(\sum_i \frac{\alpha_i\mu_i}{6}, \frac{5}{6}, \sum_i \frac{\alpha_i(1-\mu_i)}{6}\right) - \sum_i \alpha_i H\left(\frac{\mu_i}{6}, \frac{5}{6}, \frac{1-\mu_i}{6}\right)\\&= \frac{1}{6}H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i),\\I(V_1; Y_{21}) &= H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) - \sum_i \alpha_i H(\mu_i, 1-\mu_i),\\I(X_1; Y_{11}|V_1) &= \sum_i \alpha_i H\left(\frac{\mu_i}{2}, \frac{1}{2}, \frac{1-\mu_i}{2}\right) - \sum_i \alpha_i\mu_i H\left(\frac{1}{2}, \frac{1}{2}\right) - \sum_i \alpha_i(1-\mu_i) H\left(\frac{1}{2}, \frac{1}{2}\right)\\&= \frac{1}{2}\sum_i \alpha_i H(\mu_i, 1-\mu_i).\end{aligned}$$

Similarly, let $\mathrm{P}\{V_2 = i\} = \beta_i,\ \mathrm{P}\{X_2 = 0|V_2 = i\} = \nu_i$. Then

$$\begin{aligned}I(V_2; Y_{31}) &= \frac{1}{2}H\left(\sum_i \beta_i\nu_i, \sum_i \beta_i(1-\nu_i)\right) - \frac{1}{2}\sum_i \beta_i H(\nu_i, 1-\nu_i),\\I(X_2; Y_{12}|V_2) &= \sum_i \beta_i H(\nu_i, 1-\nu_i).\end{aligned}$$

Now setting $\sum_i \beta_i H(\nu_i, 1-\nu_i) = 1-q$, and $\sum_i \alpha_i H(\mu_i, 1-\mu_i) = 1-p$, we obtain

$$\begin{aligned}
I(U_1;Y_{31}) &= \frac{1}{6}H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \\
&\le \frac{1}{6}(1-(1-p)) = \frac{p}{6}, \\
I(U_1;Y_{21}) &= H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) - \sum_i \alpha_i H(\mu_i, 1-\mu_i) \\
&\le 1-(1-p) = p, \\
I(X_1;Y_{11}|U_1) &= \frac{1-p}{2}, \\
I(U_2;Y_{31}) &= \frac{1}{6}H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \\
&\le \frac{1}{2}(1-(1-q)) = \frac{q}{2}, \\
I(X_2;Y_{12}|U_2) &= 1-q.
\end{aligned}$$

Therefore, any rate pair in the BZT region must satisfy the conditions

$$\begin{aligned}
R_0 &\le \min\left\{\frac{p}{6}+\frac{q}{2}, p\right\}, \\
R_1 &\le \frac{1-p}{2} + 1 - q.
\end{aligned}$$

for some $0 \le p, q \le 1$.

It is easy to see that equality is achieved when the marginals of V_1 are given by $\mathrm{P}\{V_1 = 0\} = \mathrm{P}\{V_1 = 1\} = p/2,\ \mathrm{P}\{V_1 = E\} = 1-p$ and the marginals of V_2 are given by $\mathrm{P}\{V_2 = 0\} = \mathrm{P}\{V_2 = 1\} = q/2, \mathrm{P}\{V_2 = E\} = 1-q$, (see Figure 4).

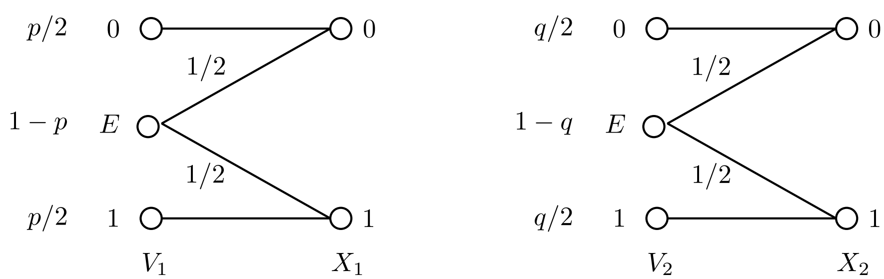

Fig. 4. Auxiliary channels that achieve the boundary of the BZT region.

Proof of Proposition 4:

We prove that the region (9) reduces to region (11) for the binary erasure channel described by the example in Section IV.

Assume that $\mathrm{P}\{V_{11} = i\} = \alpha_i, \mathrm{P}\{X_1 = 0|V_{11} = i\} = \mu_i, \mathrm{P}\{V_{12} = i\} = \beta_i, \mathrm{P}\{X_2 = 0|V_{12} = i\} =$

$\nu_i, \mathrm{P}\{V_{21} = i\} = \gamma_i, \mathrm{P}\{X_1 = 0|V_{21} = i\} = \omega_i$. Further, there exist $r, s, t \in [0, 1]$ such that

$$H(X_1|V_{11}) = \sum_i \alpha_i H(\mu_i, 1 - \mu_i) = 1 - r,$$
$$H(X_2|V_{12}) = \sum_i \beta_i H(\nu_i, 1 - \nu_i) = 1 - s,$$
$$H(X_1|V_{21}) = \sum_i \gamma_i H(\omega_i, 1 - \omega_i) = 1 - t.$$

Clearly from the Markov condition $V_{11} \to V_{21} \to X_1$, we require $1 - t \le 1 - r$ or equivalently $r \le t$.

We can also establish the following in a similar fashion.

$$I(V_{11}; Y_{31}) = \frac{1}{6} H\left(\sum_i \alpha_i \mu_i, \sum_i \alpha_i (1 - \mu_i)\right) - \frac{1}{6} \sum_i \alpha_i H(\mu_i, 1 - \mu_i) \le \frac{r}{6},$$
$$I(V_{12}; Y_{32}) = \frac{1}{2} H\left(\sum_i \beta_i \nu_i, \sum_i \beta_i (1 - \nu_i)\right) - \frac{1}{2} \sum_i \beta_i H(\nu_i, 1 - \nu_i) \le \frac{s}{2},$$
$$I(V_{21}; Y_{21}) = H\left(\sum_i \gamma_i \omega_i, \sum_i \gamma_i (1 - \omega_i)\right) - \sum_i \gamma_i H(\omega_i, 1 - \omega_i) \le t,$$
$$I(X_1; Y_{11}|V_{11}) = \frac{1}{2} \sum_i \alpha_i H(\mu_i, 1 - \mu_i) = \frac{1 - r}{2},$$
$$I(X_2; Y_{12}|V_{12}) = \sum_i \beta_i H(\nu_i, 1 - \nu_i) = 1 - s,$$
$$I(X_1; Y_{11}|V_{21}) = \frac{1}{2} \sum_i \gamma_i H(\omega_i, 1 - \omega_i) = \frac{1 - t}{2}.$$

Thus any rate pair in the capacity region must satisfy

$$R_0 \le \min\left\{\frac{r}{6} + \frac{s}{2}, t\right\},$$
$$R_0 + R_1 \le \min\left\{\frac{r}{6} + \frac{s}{2} + \frac{1 - r}{2} + 1 - s, t + \frac{1 - t}{2} + 1 - s\right\},$$

for some $0 \le r \le t \le 1, 0 \le s \le 1$. Note that substituting $r = t$ yields the BZT region.

Equality in the above conditions is achieved by the choices of auxiliary random variables shown in Figure 5, and thus the above region is the capacity region.

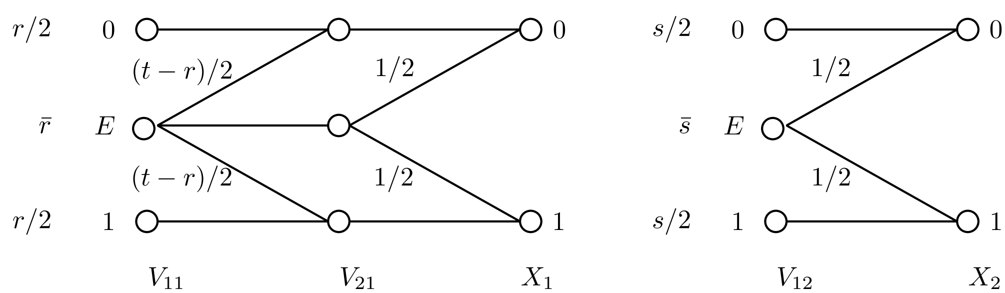

Fig. 5. Auxiliary channels that achieve the boundary of the capacity region.

APPENDIX II
FOURIER-MOTZKIN ELIMINATION FOR PROPOSITION 5

In this section we provide the details of the Fourier-Motzkin procedure in the proof of Proposition 5. To eliminate T_2, T_3 we need to consider the following set of inequalities

$$\begin{aligned}
S_2 &\leq T_2,\\
S_3 &\leq T_3,\\
S_2 + S_3 &\leq T_2 + T_3 - I(U_2;U_3|U_1),\\
R_0 + T_2 &\leq I(U_2;Y_2),\\
R_0 + T_3 &\leq I(U_2;Y_3).
\end{aligned}$$

Elimination T_2 first we end up with

$$\begin{aligned}
S_3 &\leq T_3,\\
R_0 + S_2 + S_3 &\leq I(U_2;Y_2) + T_3 - I(U_2;U_3|U_1),\\
R_0 + S_2 &\leq I(U_2;Y_2),\\
R_0 + T_3 &\leq I(U_2;Y_3).
\end{aligned}$$

Elimination of T_3 in the above leads us to

$$\begin{aligned}
2R_0 + S_2 + S_3 &\leq I(U_2;Y_2) + I(U_2;Y_3) - I(U_2;U_3|U_1),\\
R_0 + S_2 &\leq I(U_2;Y_2),\\
R_0 + S_3 &\leq I(U_2;Y_3).
\end{aligned}$$

Thus any pair $R_0, R_1 = S_1 + S_2 + S_3$ that satisfies the following set of inequalities is achievable

$$\begin{aligned}
S_1 &\geq 0,\\
S_2 &\geq 0,\\
S_3 &\geq 0,\\
R_0 + S_2 &\leq I(U_2;Y_2),\\
R_0 + S_3 &\leq I(U_3;Y_3),\\
2R_0 + S_2 + S_3 &\leq I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1),\\
R_0 + S_2 + S_3 + S_1 &\leq I(X;Y_1),\\
S_2 + S_3 + S_1 &\leq I(X;Y_1|U_1),\\
S_2 + S_1 &\leq I(X;Y_1|U_3),\\
S_3 + S_1 &\leq I(X;Y_1|U_2),\\
S_1 &\leq I(X;Y_1|U_2,U_3).
\end{aligned}$$

Substituting for $S_1 = R_1 - S_2 - S_3$ yields

$$\begin{aligned}
S_2 &\geq 0,\\
S_3 &\geq 0,\\
S_2 + S_3 &\leq R_1,\\
R_0 + S_2 &\leq I(U_2;Y_2),\\
R_0 + S_3 &\leq I(U_3;Y_3),\\
2R_0 + S_2 + S_3 &\leq I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1),\\
R_0 + R_1 &\leq I(X;Y_1),\\
R_1 &\leq I(X;Y_1|U_1),\\
R_1 &\leq S_3 + I(X;Y_1|U_3),\\
R_1 &\leq S_2 + I(X;Y_1|U_2),\\
R_1 &\leq S_2 + S_3 + I(X;Y_1|U_2,U_3).
\end{aligned}$$

Elimination of S_2 leads to

$$\begin{aligned}
0 &\leq S_3\\
R_0 + S_3 &\leq I(U_3;Y_3),\\
R_0 + R_1 &\leq I(X;Y_1),\\
R_1 &\leq I(X;Y_1|U_1),\\
R_1 &\leq S_3 + I(X;Y_1|U_3),\\
S_3 &\leq R_1,\\
R_0 &\leq I(U_2;Y_2),\\
2R_0 + S_3 &\leq I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1),\\
S_3 &\leq I(X;Y_1|U_2),\\
R_0 + R_1 &\leq I(U_2;Y_2) + I(X;Y_1|U_2),\\
2R_0 + R_1 + S_3 &\leq I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1) + I(X;Y_1|U_2),\\
0 &\leq I(X;Y_1|U_2,U_3) \text{ redundant},\\
R_0 + R_1 &\leq I(U_2;Y_2) + S_3 + I(X;Y_1|U_2,U_3),\\
2R_0 + R_1 &\leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(U_2;U_3|U_1).
\end{aligned}$$

Finally eliminating S_3 (and removing redundant inequalities) leads one to the region in Proposition 5.

APPENDIX III
PROOF OF REMARK 1 FOLLOWING THEOREM 2

Consider the 3-receiver broadcast channel with 3 degraded message sets. Let $R_2 = S_1 + S_2 + S_3$. The proof is in three steps:

(i) First, we show that any rate tuple $(R_0, R_1, S_1, S_2, S_3)$ is achievable provided

$$\begin{aligned}
R_1 &\leq T_{21}, \\
S_2 &\leq T_{22}, \\
S_3 &\leq T_3, \\
R_1 + S_3 &\leq T_{21} + T_3 - I(\tilde{U}_2; U_3|U_1), \\
R_1 + S_2 + S_3 &\leq T_{21} + T_{22} + T_3 - I(U_2; U_3|U_1), \\
R_0 + S_1 + R_1 + S_2 + S_3 &\leq I(X; Y_1), \\
S_1 + R_1 + S_2 + S_3 &\leq I(X; Y_1|U_1), \\
S_1 + S_2 + S_3 &\leq I(X; Y_1|\tilde{U}_2), \\
S_1 + S_3 &\leq I(X; Y_1|U_2), \\
S_1 + R_1 + S_2 &\leq I(X; Y_1|U_3), \\
S_1 + S_2 &\leq I(X; Y_1|U_3, \tilde{U}_2), \\
S_1 &\leq I(X; Y_1|U_2, U_3), \\
R_0 + T_{21} + T_{22} &\leq I(U_2; Y_2), \\
T_{21} &\leq I(\tilde{U}_2; Y_2|U_1), \\
R_0 + T_3 &\leq I(U_3; Y_3),
\end{aligned} \tag{32}$$

for $p(u_1, \tilde{u}_2, u_2, u_3, x) = p(u_1)p(\tilde{u}_2|u_1)p(u_2|\tilde{u}_2)p(x, u_3|u_2) = p(u_1)p(u_3|u_1)p(\tilde{u}_2, u_2|u_3)p(x|u_2, u_3)$, i.e. $U_1 \to U_3 \to (\tilde{U}_2, U_2, X)$ and $U_1 \to \tilde{U}_2 \to U_2 \to (U_3, X)$ form Markov chains.

(ii) Then, we show that the region defined by (32) is equal to the inner bound in Theorem 2.

(iii) Finally we show that when $R_1 = 0$, the conditions (32) reduce to conditions (15)-(22) in the proof of Proposition 5, thus completing the proof of Remark 1.

A. *Achievability of Rates Satisfying* (32)

First we outline the achievability of any rate tuple $(R_0, R_1, S_1, S_2, S_3)$ that satisfies conditions (32). Code generation is very similar to that in the proof of Proposition 5. We insert $\tilde{U}_2$, an auxiliary random variable representing the information about M_1, between U_1 and U_2; so for every $U_1^n(m_0)$ we generate $2^{nT_{21}}$ $\tilde{U}_2^n(m_0, m_1)$ sequences and randomly partition them into 2^{nR_1} bins. For each $\tilde{U}_2^n(m_0, m_1)$, we generate $2^{nT_{22}}$ $U_2^n(m_0, m_1, t_{21})$ sequences and randomly partition them into 2^{nS_2} bins. We then generate 2^{nT_3} $U_3^n(m_0, t_3)$ sequences and partition them into 2^{nS_3} bins. For each product bin $((m_1, s_2), s_3)$ we select a jointly typical pair $(U_2^n(m_0, m_1, t_2), U_3^n(m_0, t_3))$. Finally for product bin $((m_1, s_2), s_3)$ with corresponding jointly typical $(U_2^n(m_0, m_1, t_2), U_3^n(m_0, t_3))$ pair, we generate 2^{nS_1} sequences $X^n(m_0, m_1, s_2, s_3, s_1)$.

To ensure correct code generation (existence of relevant jointly typical sequences) we require that

$$\begin{aligned}
R_1 &\leq T_{21}, \\
S_2 &\leq T_{22}, \\
S_3 &\leq T_3, \\
R_1 + S_3 &\leq T_{21} + T_3 - I(\tilde{U}_2; U_3|U_1), \\
R_1 + S_2 + S_3 &< T_{21} + T_{22} + T_3 - I(U_2; U_3|U_1).
\end{aligned}$$

Receiver Y_1 uses joint typicality to find $(m_0, m_1, s_1, s_2, s_3)$. The following conditions on the probability of error ensure successful decoding (the corresponding events that partition the error event are listed).

$$\begin{aligned}
R_0 + S_1 + R_1 + S_2 + S_3 &< I(X;Y_1), && \text{(event: } \hat{m}_0 \neq 1)\\
S_1 + R_1 + S_2 + S_3 &< I(X;Y_1|U_1), && \text{(event: } (\hat{m}_0 = 1, \hat{m}_1 \neq 1, \hat{s}_3 \neq 1))\\
S_1 + S_2 + S_3 &< I(X;Y_1|\tilde{U}_2), && \text{(event: } (\hat{m}_0 = 1, \hat{m}_1 = 1, \hat{s}_2 \neq (1,1), \hat{s}_3 \neq 1))\\
S_1 + S_3 &< I(X;Y_1|U_2), && \text{(event: } (\hat{m}_0 = 1, \hat{m}_1 = 1, \hat{s}_2 = (1,1), \hat{s}_3 \neq 1))\\
S_1 + R_1 + S_2 &< I(X;Y_1|U_3), && \text{(event: } (\hat{m}_0 = 1, \hat{s}_3 = 1, \hat{m}_1 \neq 1))\\
S_1 + S_2 &< I(X;Y_1|U_3, \tilde{U}_2), && \text{(event: } (\hat{m}_0 = 1, \hat{s}_3 = 1, \hat{m}_1 = 1, \hat{s}_2 \neq (1,1)))\\
S_1 &< I(X;Y_1|U_2, U_3), && \text{(event: } (\hat{m}_0 = 1, \hat{s}_3 = 1, \hat{m}_1 = 1, \hat{s}_2 = (1,1), \hat{s}_1 \neq 1)).
\end{aligned}$$

Receiver Y_2 decodes m_0 via indirect decoding using U_2 and m_1 by decoding $\tilde{U}_2$ conditioned on U_1. This is successful provided

$$\begin{aligned}
R_0 + T_{21} + T_{22} &< I(U_2;Y_2),\\
T_{21} &< I(\tilde{U}_2;Y_2|U_1).
\end{aligned}$$

Receiver Y_3 decodes m_0 via indirect decoding using U_3. This step succeeds provided

$$R_0 + T_3 < I(U_3;Y_3).$$

Combining the above conditions we see that any rate tuple satisfying (32) is achievable.

B. *Equivalence of Conditions* (32) *to Theorem 2*

In one direction, setting $\tilde{U}_2 = U_2, S_2 = 0$, $T_{22} = 0$ and $T_{21} = T_2$, we obtain (28). Thus conditions (32) contain the region described by Theorem 2.

For the reverse direction we break down the argument into two cases.

Case 1: $T_{22} < I(U_2;Y_2|\tilde{U}_2)$

Observe that Y_2 can also decode S_2 and setting $\tilde{R}_1 = R_1 + S_2$, $\tilde{R}_2 = R_2 - S_2$, and $T_{21} + T_{22} = T_2$ we see that conditions (32) along with $T_{22} < I(U_2;Y_2|\tilde{U}_2)$ imply conditions (28). Thus under $T_{22} < I(U_2;Y_2|\tilde{U}_2)$, the region described by (32) is contained in the region described by Theorem 2.

Case 2: $T_{22} < I(U_2;Y_2|\tilde{U}_2)$

If $T_{22} \geq I(U_2;Y_2|\tilde{U}_2)$, then the condition $R_0 + T_{21} + T_{22} < I(U_2;Y_2)$ implies that $R_0 + T_{21} < I(\tilde{U}_2;Y_2)$ and Y_2's requirement for successful decoding can be changed to

$$\begin{aligned}
R_0 + T_{21} &< I(\tilde{U}_2;Y_2),\\
T_{21} &< I(\tilde{U}_2;Y_2|U_1).
\end{aligned}$$

In the rest of the inequalities, replacing U_2 by $\tilde{U}_2$ only weakens them and hence it is optimal to set $U_2 = \tilde{U}_2$. These new inequalities imply (28) in which we replace S_1 by $S_1 + S_2$ and U_2 by $\tilde{U}_2$. Thus under $T_{22} > I(U_2;Y_2|\tilde{U}_2)$ also, the region described by (32) is contained in the region described by Theorem 2.

Combining Cases 1 and 2 we see that rate pairs satisfying conditions (32) is contained in the region described by Theorem 2. This completes the proof of their equivalence.

C. Reduction to Proposition 5

If $R_1 = 0$, the region described by conditions (32) reduce to

$$\begin{aligned}
0 &\le T_{21}, \\
S_2 &\le T_{22}, \\
S_3 &\le T_3, \\
S_3 &\le T_{21} + T_3 - I(\tilde{U}_2; U_3|U_1), \\
S_2 + S_3 &\le T_{21} + T_{22} + T_3 - I(U_2; U_3|U_1), \\
S_1 + S_2 + S_3 &\le I(X; Y_1|\tilde{U}_2), \\
R_0 + S_1 + S_2 + S_3 &\le I(X; Y_1), \\
S_1 + S_2 + S_3 &\le I(X; Y_1|U_1), \\
S_1 + S_3 &\le I(X; Y_1|U_2), \\
S_1 + S_2 &\le I(X; Y_1|U_3), \\
S_1 + S_2 &\le I(X; Y_1|U_3, \tilde{U}_2) \\
S_1 &\le I(X; Y_1|U_2, U_3) \\
R_0 + T_{21} + T_{22} &\le I(U_2; Y_2), \\
T_{21} &\le I(\tilde{U}_2; Y_2|U_1) \\
R_0 + T_3 &\le I(U_3; Y_3).
\end{aligned} \tag{33}$$

Recalling that $R_2 = S_2 + S_3 + S_1$, and setting $\tilde{T}_{22} = T_{21} + T_{22}$, observe that any (R_0, S_1, S_2, S_3) satisfying the above inequalities (33) also satisfies

$$\begin{aligned}
S_2 &\le \tilde{T}_{22}, \\
S_3 &\le T_3, \\
S_2 + S_3 &\le \tilde{T}_{22} + T_3 - I(U_2; U_3|U_1), \\
R_0 + S_1 + S_2 + S_3 &\le I(X; Y_1), \\
S_1 + S_2 + S_3 &\le I(X; Y_1|\tilde{U}_2), \\
S_1 + S_3 &\le I(X; Y_1|U_2), \\
S_1 + S_2 &\le I(X; Y_1|U_3), \\
S_1 + S_2 &\le I(X; Y_1|U_3, \tilde{U}_2), \\
S_1 &\le I(X; Y_1|U_2, U_3), \\
R_0 + \tilde{T}_{22} &\le I(U_2; Y_2), \\
R_0 + T_3 &\le I(U_3; Y_3).
\end{aligned}$$

These conditions are clearly maximized by setting $\tilde{U}_2 = U_1$ which in turn reduces the equations to conditions (15)-(22) of 5.. Thus the region defined by (33) is contained in the region given by Proposition 5. The other direction is direct as the region in Proposition 5 is obtained by setting $\tilde{U}_2 = U_1$ in (33). This completes the proof of Remark 1.

Remark: Observe that we do not need the auxiliary random variable $\tilde{U}_2$ to characterize the region in either the 3 degraded message sets case (Theorem 2) or the 2 degraded message sets case (Proposition 5). This is in accordance with the structure of auxiliary random variables as prescribed by the remark in the introduction of Subsection V-B.